Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 98 (2007) 1726 – 1750
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Asymptotic expansions of the distributions of estimators in canonical correlation analysis under nonnormality☆

Haruhiko Ogasawara*

Department of Information and Management Science, Otaru University of Commerce, 3-5-21 Midori, Otaru 047-8501, Japan

Received 1 December 2005
Available online 17 January 2007

Abstract

Asymptotic expansions of the distributions of typical estimators in canonical correlation analysis under nonnormality are obtained. The expansions include the Edgeworth expansions up to order $O(1/n)$ for the parameter estimators standardized by the population standard errors, and the corresponding expansion by Hall's method with variable transformation. The expansions for the Studentized estimators are also given using the Cornish–Fisher expansion and Hall's method. The parameter estimators are dealt with in the context of estimation for the covariance structure in canonical correlation analysis. The distributions of the associated statistics (the structure of the canonical variables, the scaled log likelihood ratio and Rozeboom's between-set correlation) are also expanded. The robustness of the normal-theory asymptotic variances of the sample canonical correlations and associated statistics are shown when a latent variable model holds. Simulations are performed to see the accuracy of the asymptotic results in finite samples.

AMS 1991 subject classification: 62E20; 62H20

Keywords: Edgeworth expansion; Variable transformation; Studentized estimators; Nonnormality; Covariance structure; Asymptotic robustness

☆ Partially supported by Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

* Fax: +81 134 27 5376.
E-mail address: hogasa@res.otaru-uc.ac.jp.

0047-259X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2006.12.001

1. Introduction

Canonical correlation analysis was initiated by Hotelling [17]. The asymptotic standard errors of the sample canonical correlations under normality were provided by Hotelling [18, Eq. (5.27)]. Bartlett [5] derived the asymptotic chi-square null distribution of the likelihood ratio statistic under normality. Hsu [19] gave the asymptotic distributions of the sample canonical correlations under normality though the expression is somewhat intractable for actual computation. Under the same condition, Lawley [21] provided the asymptotic cumulants of the distributions of the sample canonical correlations up to the fourth order. In his derivation, however, the higher-order asymptotic variances or the added asymptotic variances up to order $O(n^{-2})$ were given only as approximations in some special cases since "the term of order n^{-2} is extremely cumbersome" (p. 61), where $n + 1$ is the sample size.

Anderson [1, Section 13.4] gave the exact distributions of the squared sample canonical correlations when two sets of variables are uncorrelated under normality. This result was generalized by Constantine [9] to the correlated cases though the expression is involved due to the use of a hyper-geometric function. For the likelihood ratio statistic with the assumption of normality, Fujikoshi [13] derived its asymptotic distributions under local and fixed alternative hypotheses. Anderson [2] gave the asymptotic covariance matrix of the sample canonical correlations and the corresponding sample coefficients of the canonical variables under normality.

The normality assumption used in the above articles has been relaxed in various aspects. Muirhead and Waternaux [23] derived the asymptotic covariance matrix of the squared sample canonical correlations, and the asymptotic distributions of the likelihood ratio statistic under local and fixed alternatives with distinct population roots. Under the similar condition, Steiger and Browne [30] gave the asymptotic variances of the sample canonical correlations using the results for usual sample correlation coefficients.

Fang and Krishnaiah [12] gave the singe-term Edgeworth expansion of the distribution of the squared sample canonical correlation up to order $O(n^{-1/2})$ under nonnormality in the cases with possibly multiple population roots. For elliptically distributed cases with multiple roots, Eaton and Tyler [11] provided the asymptotic covariance matrix for the functions of the squared sample canonical correlations. Boik [6] derived the asymptotic covariance matrix and asymptotic biases of the sample canonical correlations and the sample coefficients of the canonical variables under nonnormality including multiple roots. Recently, Bai and He [4] gave the necessary and sufficient condition of the robustness of the normal-theory (NT) asymptotic distribution of the likelihood ratio statistic against the violation of the normality assumption in the cases including multiple roots.

General explanation about canonical correlation analysis is found in e.g., Siotani et al. [29, Chapter 22], Rencher [27, Chapter 11] and Anderson [3, Chapters 12 and 13]. The purpose of this study is to give the Edgeworth expansions of the various estimators with standardization using population asymptotic standard errors including the likelihood ratio statistic in canonical correlation analysis up to order $O(n^{-1})$ under nonnormality. The asymptotic expansions of the distributions of the Studentized parameter estimators will also be given under normality and nonnormality with the Cornish–Fisher expansion and Hall's [15] method by variable transformation. Simulations will be performed to see the accuracy of our formulas in finite samples.

2. The covariance structure in canonical correlation analysis

In this article, canonical correlation analysis is dealt with in the context of covariance structure analysis including canonical correlations and associated coefficients as parameters. Let **x** and **y** be the $p \times 1$ and $q \times 1$ vectors of observable random variables with $p \leqslant q$, respectively. The $(p+q) \times (p+q)$ covariance matrix of $\mathbf{x}$ and $\mathbf{y}$ is denoted by

$$\mathrm{Cov}\{(\mathbf{x}', \mathbf{y}')'\} = \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{YX} & \mathbf{\Sigma}_{YY} \end{bmatrix}, \tag{2.1}$$

with the corresponding unbiased sample covariance matrix based on $N = n + 1$ observations being

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{bmatrix}. \tag{2.2}$$

Two sets of p canonical variables, under the standard definition, are denoted by the $p \times 1$ vectors $\mathbf{f}$ and $\mathbf{g}$ with

$$\mathbf{f}' = (\mathbf{x} - \boldsymbol{\mu}_X)'\mathbf{A} \quad \text{and} \quad \mathbf{g}' = (\mathbf{y} - \boldsymbol{\mu}_Y)'\mathbf{B}_1, \tag{2.3}$$

where $\boldsymbol{\mu}_X = \mathrm{E}(\mathbf{x})$ and $\boldsymbol{\mu}_Y = \mathrm{E}(\mathbf{y})$; $\mathbf{A}$ and $\mathbf{B}_1$ are the $p \times p$ and $q \times p$ matrices of the coefficients of canonical variables, respectively;

$$\begin{aligned} &\mathrm{Cov}(\mathbf{f}) = \mathbf{A}'\mathbf{\Sigma}_{XX}\mathbf{A} = \mathbf{I}_p, \quad \mathrm{Cov}(\mathbf{g}) = \mathbf{B}_1'\mathbf{\Sigma}_{YY}\mathbf{B}_1 = \mathbf{I}_p, \\ &\mathrm{Cov}\{(\mathbf{f}', \mathbf{g}')'\} = \begin{bmatrix} \mathbf{I}_p & \mathbf{\Lambda} \\ \mathbf{\Lambda} & \mathbf{I}_p \end{bmatrix}, \quad \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) \equiv \mathrm{diag}(\boldsymbol{\lambda}'), \end{aligned} \tag{2.4}$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix; and the canonical correlations λ_i are assumed to be $1 > \lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$ (the case with multiple roots will be addressed in the last section).

When $p < q$, let

$$\mathbf{B} = [\mathbf{B}_1 \quad \mathbf{B}_2] = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{B}^{-1} = \begin{bmatrix} \mathbf{B}^{11} & \mathbf{B}^{12} \\ \mathbf{B}^{21} & \mathbf{B}^{22} \end{bmatrix}, \tag{2.5}$$

with the assumption of the existence of the inverse of $\mathbf{B}$ (and the inverses of matrices in the following), where $\mathbf{B}_2$ is the $q \times (q-p)$ matrix of the coefficients of residual variables for $\mathbf{y}$ with $\mathbf{B}'\mathbf{\Sigma}_{YY}\mathbf{B} = \mathbf{I}_q$.

In the remaining part of this article, we deal with the case of $p < q$. When $p = q$, $\mathbf{B}$ is defined as $\mathbf{B}_1$ and $\mathbf{B}_2$ with associated results can be omitted. From the above definitions, we have

$$\begin{bmatrix} \mathbf{A}' & \mathbf{O} \\ \mathbf{O} & \mathbf{B}' \end{bmatrix} \mathbf{\Sigma} \begin{bmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_p & \mathbf{\Lambda} & \mathbf{O} \\ \mathbf{\Lambda} & \mathbf{I}_p & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{I}_{q-p} \end{bmatrix}$$

or

$$\mathbf{\Sigma} = \begin{bmatrix} (\mathbf{A}\mathbf{A}')^{-1} & \mathbf{A}'^{-1}\mathbf{\Lambda}[\mathbf{B}^{11} \quad \mathbf{B}^{12}] \\ \begin{bmatrix} \mathbf{B}'^{11} \\ \mathbf{B}'^{12} \end{bmatrix} \mathbf{\Lambda}\mathbf{A}^{-1} & (\mathbf{B}\mathbf{B}')^{-1} \end{bmatrix}. \tag{2.6}$$

(Compare the usual decomposition of $\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}$ whose sample counterpart is addressed in the next section.) Note that the second equation of (2.6) shows the covariance structure in canonical correlation analysis. The $(p^2 + q^2 + p) \times 1$ vector $\boldsymbol{\theta}$ of population parameters in (2.6) can be defined as

$$\boldsymbol{\theta} = (\mathrm{vec}'\mathbf{A}, \mathrm{vec}'\mathbf{B}, \boldsymbol{\lambda}')' \quad \text{or} \quad \boldsymbol{\theta} = (\mathrm{vec}'\mathbf{A}'^{-1}, \mathrm{vec}'\mathbf{B}'^{-1}, \boldsymbol{\lambda}')', \tag{2.7}$$

where $\mathrm{vec}(\cdot)$ is the vectorizing operator stacking the columns of an argument matrix with $\mathrm{vec}'(\cdot) = \{\mathrm{vec}(\cdot)\}'$. We employ the latter definition of $\boldsymbol{\theta}$ in (2.7) since $\boldsymbol{\Sigma}$ is quadratic or multi-linear with respect to the parameters by this formulation, and since $\mathbf{A}'^{-1}$ and $\mathbf{B}'^{-1}$ have meanings as structure-covariances or the covariances between canonical variables and associated observable variables i.e., $\mathbf{A}'^{-1} = \boldsymbol{\Sigma}_{XX}\mathbf{A}$ and $\mathbf{B}'^{-1} = \boldsymbol{\Sigma}_{YY}\mathbf{B}$ (see (2.4) and (2.5)). The coefficient matrices $\mathbf{A}$ and $\mathbf{B}$ will be treated as transformed parameters afterwards.

The covariance structure of (2.6) has the rotational indeterminacy for the rows of $[\mathbf{B}^{21}\ \ \mathbf{B}^{22}]$ when $q \geqslant p+2$ in addition to the minor indeterminacy of the signs of the rows of $[\mathbf{A}^{-1}\ \ \mathbf{B}^{11}\ \ \mathbf{B}^{12}]$ and $[\mathbf{B}^{21}\ \ \mathbf{B}^{22}]$. The indeterminacy of $[\mathbf{B}^{21}\ \ \mathbf{B}^{22}]$ can be removed by using orthogonal rotation as in factor analysis. However, the matrix $[\mathbf{B}^{21}\ \ \mathbf{B}^{22}]$ or $\mathbf{B}_2$ is usually the nuisance parameters of little interest. So, one of the convenient methods, employed in this article, is to fix appropriate $\{(q-p)^2 - (q-p)\}/2$ elements in $[\mathbf{B}^{21}\ \ \mathbf{B}^{22}]$ at 0 to have e.g., an echelon form of $[\mathbf{B}^{21}\ \ \mathbf{B}^{22}]$ (note that $\mathbf{B}_1$ is unchanged by the rotation of $[\mathbf{B}^{21}\ \ \mathbf{B}^{22}]$). For this formulation, $\boldsymbol{\theta}$ is redefined as a $Q \times 1$ vector consisting of $\mathrm{vec}(\mathbf{A}'^{-1})$, the nonfixed elements in $\mathrm{vec}(\mathbf{B}'^{-1})$ and $\boldsymbol{\lambda}$, where $Q = p^2 + q^2 + p - \{(q-p)^2 - (q-p)\}/2 = \{(p+q)^2 + (p+q)\}/2$, which is equal to the number of the nonduplicated elements in $\boldsymbol{\Sigma}$.

3. The estimators of parameters and their asymptotic distributions

The vector $\hat{\boldsymbol{\theta}}$ of parameter estimators based on $\mathbf{S}$ is usually given from the spectral decomposition of $\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1/2}$, whose eigenvalues are $\hat{\lambda}_i^2$ $(i = 1, \ldots, p)$ with the corresponding eigenvectors being the columns of $\mathbf{S}_{XX}^{1/2}\hat{\mathbf{A}}$, where $(\mathbf{S}_{XX}^{1/2})^2 = \mathbf{S}_{XX}$ and $(\mathbf{S}_{XX}^{-1/2})^2 = \mathbf{S}_{XX}^{-1}$. The matrix $\hat{\mathbf{B}}$ is obtained by $\hat{\mathbf{B}}_1 = \mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\hat{\mathbf{A}}\hat{\boldsymbol{\Lambda}}^{-1}$ with $\hat{\boldsymbol{\Lambda}} = \mathrm{diag}(\hat{\boldsymbol{\lambda}}') = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ and $\hat{\mathbf{B}}_2\hat{\mathbf{B}}_2' = \mathbf{S}_{YY}^{-1} - \hat{\mathbf{B}}_1\hat{\mathbf{B}}_1'$. The estimators of structure covariances are given by $\mathbf{S}_{XX}\hat{\mathbf{A}} = \hat{\mathbf{A}}'^{-1}$ and $\mathbf{S}_{YY}\hat{\mathbf{B}} = \hat{\mathbf{B}}'^{-1}$.

The vector $\hat{\boldsymbol{\theta}}$ is seen as a function $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\mathbf{s})$ of the $Q \times 1$ vector $\mathbf{s} = \mathrm{v}(\mathbf{S})$, where $\mathrm{v}(\cdot)$ is the vectorizing operator taking the nonduplicated elements of a symmetric matrix, though the function $\boldsymbol{\theta}(\cdot)$ is an implicit one given from $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = \mathbf{S}$. Let $\hat{\theta}$ be an element of $\hat{\boldsymbol{\theta}}$. Then, $\hat{\theta}$ is assumed to be expressed by the Taylor series as

$$
\begin{aligned}
\hat{\theta} = \theta &+ \left.\frac{\partial\hat{\theta}}{\partial\mathbf{s}'}\right|_{\mathbf{s}=\boldsymbol{\sigma}} (\mathbf{s}-\boldsymbol{\sigma}) + \frac{1}{2}\left(\frac{\partial}{\partial\mathbf{s}'}\right)^{\langle 2\rangle} \hat{\theta}|_{\mathbf{s}=\boldsymbol{\sigma}}(\mathbf{s}-\boldsymbol{\sigma})^{\langle 2\rangle} \\
&+ \frac{1}{6}\left(\frac{\partial}{\partial\mathbf{s}'}\right)^{\langle 3\rangle} \hat{\theta}|_{\mathbf{s}=\boldsymbol{\sigma}}(\mathbf{s}-\boldsymbol{\sigma})^{\langle 3\rangle} + o_p(n^{-3/2}),
\end{aligned} \tag{3.1}
$$

where θ is the population value of $\hat{\theta}$; $\boldsymbol{\sigma} = \mathrm{v}(\boldsymbol{\Sigma})$; $\mathbf{X}^{\langle k\rangle} = \mathbf{X}\otimes\cdots\otimes\mathbf{X}$ (k times); $\otimes$ denotes Kronecker product; and $\mathbf{s}$ is also used as a mathematical vector variable in differentiation for simplicity of notation. Let $w = n^{1/2}(\hat{\theta} - \theta)$. It is assumed that the cumulants of w and the corresponding asymptotic cumulants exist as follows:

$$
\begin{aligned}
\kappa_1(w) &= E(w) = n^{-1/2}\alpha_1 + o(n^{-1/2}), \\
\kappa_2(w) &= E[\{w - E(w)\}^2] = \alpha_2 + n^{-1}\Delta\alpha_2 + o(n^{-1}), \\
\kappa_3(w) &= E[\{w - E(w)\}^3] = n^{-1/2}\alpha_3 + o(n^{-1/2}), \\
\kappa_4(w) &= E[\{w - E(w)\}^4] - 3\{\kappa_2(w)\}^2 = n^{-1}\alpha_4 + o(n^{-1}).
\end{aligned} \tag{3.2}
$$

Then, it is known [24] that

$$\alpha_1 = \frac{1}{2}\,\mathrm{tr}\left(\frac{\partial^2\theta}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\right), \quad \alpha_2 = \frac{\partial\theta}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\theta}{\partial\boldsymbol{\sigma}},$$

$$\alpha_3 = \sum_{a\geqslant b}\sum_{c\geqslant d}\sum_{e\geqslant f}\frac{\partial\theta}{\partial\sigma_{ab}}\frac{\partial\theta}{\partial\sigma_{cd}}\frac{\partial\theta}{\partial\sigma_{ef}}(\sigma_{abcdef} - 3\sigma_{ab}\sigma_{cdef}$$
$$-6\sigma_{abc}\sigma_{def} + 2\sigma_{ab}\sigma_{cd}\sigma_{ef}) + 3\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial^2\theta}{\partial\boldsymbol{\sigma}\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\theta}{\partial\boldsymbol{\sigma}}, \tag{3.3}$$

where $n^{-1}\boldsymbol{\Omega}$ is the asymptotic covariance matrix of $\mathbf{s}$ i.e., $(n\,\mathrm{acov}(\mathbf{s}))_{ab,cd} = (\boldsymbol{\Omega})_{ab,cd} = \sigma_{abcd} - \sigma_{ab}\sigma_{cd}$ with $\mathrm{acov}(\cdot)$ being the asymptotic covariance matrix of order $O(n^{-1})$ for the argument vector; $(\cdot)_{ab,cd}$ denotes the (a,b)th row and the (c,d)th column of $\boldsymbol{\Omega}$ using double subscript notation; $\partial\theta/\partial\boldsymbol{\sigma} = \partial\hat{\theta}/\partial\mathbf{s}|_{\mathbf{s}=\boldsymbol{\sigma}}$ with the similar expressions for partial derivatives for the simplicity of notation; $\sigma_{ab\ldots f} = E[\{X_a - E(X_a)\}\{X_b - E(X_b)\}\ldots\{X_f - E(X_f)\}]$ with $\sigma_{ab} = (\boldsymbol{\Sigma})_{ab}$; and $(\cdot)_{ab}$ denotes the (a,b)th element of an argument matrix.

Similar expressions of $\Delta\alpha_2$ and α_4 are available (see [24,14]) though they are not repeated here since they are somewhat involved. For $\Delta\alpha_2$ and α_4, the partial derivatives of $\hat{\boldsymbol{\theta}}$ with respect to $\mathbf{s}$ up to the third order are required. The asymptotic cumulants of observable variables up to the sixth and eighth orders are also required for $\Delta\alpha_2$ and α_4, respectively.

Using the asymptotic cumulants in (3.2), the Edgeworth expansion of the distribution function of standardized w or $\hat{\theta}$ is given with Cramér's condition for the validity as follows (see e.g., [16, Theorem 2.2; 14, p. 46]):

$$\Pr\left(\frac{w}{\alpha_2^{1/2}} \leqslant z\right) = \Phi(z) - n^{-1/2}\left\{\frac{\alpha_1}{\alpha_2^{1/2}} + \frac{\alpha_3}{6\alpha_2^{3/2}}(z^2-1)\right\}\phi(z) - n^{-1}\left\{\frac{1}{2}(\Delta\alpha_2 + \alpha_1^2)\frac{z}{\alpha_2}\right.$$
$$\left. + \left(\frac{\alpha_4}{24} + \frac{\alpha_1\alpha_3}{6}\right)\frac{z^3-3z}{\alpha_2^2} + \frac{\alpha_3^2(z^5 - 10z^3 + 15z)}{72\alpha_2^3}\right\}\phi(z) + o(n^{-1}), \tag{3.4}$$

where $\phi(z) = (1/\sqrt{2\pi})\exp(-z^2/2)$ and $\Phi(z) = \int_{-\infty}^{z}\phi(t)\,dt$.

It is known that the distribution functions given by the Edgeworth expansions up to $O(n^{-1/2})$ and $O(n^{-1})$ are not necessarily nondecreasing in finite samples. These anomalous phenomena can be avoided by using Hall's [15] method removing asymptotic skewness with monotone transformation. The asymptotic distribution and density functions by this method up to order $O(n^{-1/2})$ are

$$\Pr\left(\frac{w}{\alpha_2^{1/2}} \leqslant z\right) = \Phi\{g(z)\} + o(n^{-1/2}), \tag{3.5}$$

and

$$f\left(\frac{w}{\alpha_2^{1/2}} = z\right) = \phi\{g(z)\}\left\{\frac{n^{-1/2}\alpha_3}{6\alpha_2^{1/2}}\left(z - \frac{n^{-1/2}\alpha_1}{\alpha_2^{1/2}}\right) - 1\right\}^2 + o(n^{-1/2}), \tag{3.6}$$

respectively, with

$$g(z)=\frac{2n^{1/2}\alpha_2^{3/2}}{\alpha_3}\left[\left\{\frac{n^{-1/2}\alpha_3}{6\alpha_2^{3/2}}\left(z-\frac{n^{-1/2}\alpha_1}{\alpha_2^{1/2}}\right)-1\right\}^3-1\right]. \tag{3.7}$$

It is known that Hall's method is seen as a general saddlepoint method (see [32, Section 6]) first given by Eaton and Ronchetti [10].

The asymptotic expansions given above were derived by using population asymptotic cumulants, which are in practice unavailable. On the other hand, we have the ADF (asymptotically distribution free) Studentized estimator

$$t=\frac{n^{1/2}(\hat{\theta}-\theta)}{\hat{\alpha}_2^{1/2}}=\frac{w}{\hat{\alpha}_2^{1/2}}, \tag{3.8}$$

where $\hat{\alpha}_2$ is a consistent estimator of α_2. The asymptotic cumulants of t are

$$\kappa_1(t)=n^{-1/2}\alpha_1'+o(n^{-1/2}),\quad \kappa_2(t)=1+o(1),\quad \kappa_3(t)=n^{-1/2}\alpha_3'+o(n^{-1/2}), \tag{3.9}$$

where

$$\alpha_1'=\alpha_2^{-1/2}\alpha_1-\frac{1}{2}\alpha_2^{-3/2}\left\{\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\alpha_2}{\partial\boldsymbol{\sigma}}+\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}n\,\mathrm{acov}(\mathbf{s},\mathbf{s}_{(4)}')\frac{\partial\alpha_2}{\partial\boldsymbol{\sigma}_{(4)}}\right\},$$

$$\alpha_3'=\alpha_2^{-3/2}\alpha_3-3\alpha_2^{-3/2}\left\{\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\alpha_2}{\partial\boldsymbol{\sigma}}+\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}n\,\mathrm{acov}(\mathbf{s},\mathbf{s}_{(4)}')\frac{\partial\alpha_2}{\partial\boldsymbol{\sigma}_{(4)}}\right\}; \tag{3.10}$$

$\mathrm{acov}(\mathbf{s},\mathbf{s}_{(4)}')$ is the asymptotic cross covariance matrix of $\mathbf{s}$ and $\mathbf{s}_{(4)}'$ up to order $O(n^{-1})$; $\mathbf{s}_{(4)}$ is the $\binom{p+q+3}{4}\times 1$ vector of the nonduplicated sample fourth-order central moments of $p+q$ observable variables; and $\boldsymbol{\sigma}_{(4)}$ is its population counterpart [25].

Let $z_{\tilde{\alpha}}=\Phi^{-1}(1-\tilde{\alpha})$ (e.g., $\tilde{\alpha}=.05$). Then, the confidence interval for θ with the asymptotic confidence coefficient $1-\tilde{\alpha}$ accurate up to order $O(n^{-1/2})$ by the usual Cornish–Fisher expansion using consistent estimators $\hat{\alpha}_1'$ and $\hat{\alpha}_3'$ is

$$\hat{\theta}+[\pm z_{\tilde{\alpha}/2}-n^{-1/2}\{\hat{\alpha}_1'+(\hat{\alpha}_3'/6)(z_{\tilde{\alpha}/2}^2-1)\}]n^{-1/2}\hat{\alpha}_2^{1/2}. \tag{3.11}$$

The corresponding confidence interval given by Hall's method is

$$\hat{\theta}-n^{-1}\hat{\alpha}_2^{1/2}\hat{\alpha}_1'+6\hat{\alpha}_2^{1/2}(\hat{\alpha}_3')^{-1}[\{1-(1/2)\hat{\alpha}_3'(\pm n^{-1/2}z_{\tilde{\alpha}/2}-(n^{-1}/6)\hat{\alpha}_3')\}^{1/3}-1], \tag{3.12}$$

where the validity of the use of the sample asymptotic cumulants is shown by Hall [16, pp. 122–123].

In practice, the sample counterparts of (3.10) under nonnormality tend to be unstable since they include the sample moments up to the sixth order. On the other hand, (3.10) under normality reduces to

$$\alpha_{\mathrm{NT1}}'=\alpha_{\mathrm{NT2}}^{-1/2}\alpha_{\mathrm{NT1}}-\frac{1}{2}\alpha_{\mathrm{NT2}}^{-3/2}\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}_{\mathrm{NT}}\frac{\partial\alpha_{\mathrm{NT2}}}{\partial\boldsymbol{\sigma}},$$

$$\alpha_{\mathrm{NT3}}'=\alpha_{\mathrm{NT2}}^{-3/2}\alpha_{\mathrm{NT3}}-3\alpha_{\mathrm{NT2}}^{-3/2}\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}_{\mathrm{NT}}\frac{\partial\alpha_{\mathrm{NT2}}}{\partial\boldsymbol{\sigma}}, \tag{3.13}$$

where $(\boldsymbol{\Omega}_{\mathrm{NT}})_{ab,cd} = \sigma_{ac}\sigma_{bd} + \sigma_{ad}\sigma_{bc}$; and the subscript "NT" indicates the normal-theory value or the value given under normality. A typical situation encountered in practice is to misuse the NT Studentized estimator under nonnormality. The asymptotic cumulants of the statistic under such a condition is available [25] as follows:

$$\alpha''_{\mathrm{NT1}} = \alpha_{\mathrm{NT2}}^{-1/2}\alpha_1 - \frac{1}{2}\alpha_{\mathrm{NT2}}^{-3/2}\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\alpha_{\mathrm{NT2}}}{\partial\boldsymbol{\sigma}}, \quad \alpha''_{\mathrm{NT2}} = \alpha_{\mathrm{NT2}}^{-1}\alpha_2,$$

$$\alpha''_{\mathrm{NT3}} = \alpha_{\mathrm{NT2}}^{-3/2}\alpha_3 - 3\alpha_{\mathrm{NT2}}^{-5/2}\alpha_2\frac{\partial\theta}{\partial\boldsymbol{\sigma}'}\boldsymbol{\Omega}\frac{\partial\alpha_{\mathrm{NT2}}}{\partial\boldsymbol{\sigma}}. \tag{3.14}$$

Eq. (3.14) can be used to study properties of the NT Studentized estimators under nonnormality.

4. The partial derivatives of the estimators

The remaining work for actual computation in the previous section is to have the partial derivatives of $\hat{\boldsymbol{\theta}}$ with respect to $\mathbf{s}$ up to the third order. Recall that $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = \mathbf{S}$, which gives

$$\mathrm{v}(\hat{\boldsymbol{\Sigma}}) = \hat{\boldsymbol{\sigma}} = \mathbf{s}. \tag{4.1}$$

Differentiating (4.1) with respect to $\mathbf{s}$, we have

$$\frac{\partial\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'}\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'} = \frac{\partial\mathbf{s}}{\partial\mathbf{s}'} = \mathbf{I}_{P^*} \quad \text{with } P^* = (p+q)(p+q+1)/2. \tag{4.2}$$

From (4.2) with $\hat{\boldsymbol{\sigma}}$ being a one-to-one function of $\mathbf{s}$,

$$\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'} = \left(\frac{\partial\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'}\right)^{-1}. \tag{4.3}$$

The second and third partial derivatives are obtained from (4.3) recursively as

$$\begin{aligned}\frac{\partial^2\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'\partial s_{ab}} &= \frac{\partial}{\partial s_{ab}}\left(\frac{\partial\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'}\right)^{-1} = -\left(\frac{\partial\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'}\right)^{-1}\sum_{i=1}^{P^*}\frac{\partial^2\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'\partial\hat{\theta}_i}\frac{\partial\hat{\theta}_i}{\partial s_{ab}}\left(\frac{\partial\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'}\right)^{-1}\\ &= -\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'}\sum_{i=1}^{P^*}\frac{\partial^2\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'\partial\hat{\theta}_i}\frac{\partial\hat{\theta}_i}{\partial s_{ab}}\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'}\end{aligned} \tag{4.4}$$

and

$$\begin{aligned}\frac{\partial^3\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'\partial s_{ab}\partial s_{cd}} = &-\frac{\partial^2\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'\partial s_{cd}}\sum_{i=1}^{P^*}\frac{\partial^2\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'\partial\hat{\theta}_i}\frac{\partial\hat{\theta}_i}{\partial s_{ab}}\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'} - \frac{\partial\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'}\sum_{i=1}^{P^*}\left\{\sum_{j=1}^{P^*}\frac{\partial^3\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'\partial\hat{\theta}_i\partial\hat{\theta}_j}\right.\\ &\left.\times\frac{\partial\hat{\theta}_i}{\partial s_{ab}}\frac{\partial\hat{\theta}_j}{\partial s_{cd}}\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'} + \frac{\partial^2\hat{\boldsymbol{\sigma}}}{\partial\hat{\boldsymbol{\theta}}'\partial\hat{\theta}_i}\left(\frac{\partial^2\hat{\theta}_i}{\partial s_{ab}\partial s_{cd}}\frac{\partial\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'} + \frac{\partial\hat{\theta}_i}{\partial s_{ab}}\frac{\partial^2\hat{\boldsymbol{\theta}}}{\partial\mathbf{s}'\partial s_{cd}}\right)\right\}\\ &(p+q \geqslant a \geqslant b \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1).\end{aligned} \tag{4.5}$$

In (4.3), the first partial derivatives of $\hat{\mathbf{\Sigma}}$ with respect to $\hat{\boldsymbol{\theta}}$ evaluated at the population values are given using (2.6) as

$$\frac{\partial \mathbf{\Sigma}}{\partial a^{ij}} = \begin{bmatrix} \mathbf{E}_{ji}[\mathbf{A}^{-1} \;\; \mathbf{\Lambda B}^{11} \;\; \mathbf{\Lambda B}^{12}] \\ \mathbf{O} \end{bmatrix} + \left[\begin{bmatrix} \mathbf{A}'^{-1} \\ \mathbf{B}'^{11}\mathbf{\Lambda} \\ \mathbf{B}'^{12}\mathbf{\Lambda} \end{bmatrix} \mathbf{E}_{ij} \;\; \mathbf{O} \right] \quad (i, j = 1, \ldots, p),$$

$$\frac{\partial \mathbf{\Sigma}}{\partial b^{ij}} = \begin{bmatrix} \mathbf{O} & \mathbf{A}'^{-1}\mathbf{\Lambda E}_{ij} \\ \mathbf{E}_{ji}\mathbf{\Lambda A}^{-1} & \begin{bmatrix} \mathbf{B}'^{11} \\ \mathbf{B}'^{12} \end{bmatrix} \mathbf{E}_{ij} + \mathbf{E}_{ji}[\mathbf{B}^{11} \;\; \mathbf{B}^{12}] \end{bmatrix} \quad (i = 1, \ldots, p;\ j = 1, \ldots, q),$$

$$\frac{\partial \mathbf{\Sigma}}{\partial b^{p+i,j}} = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \begin{bmatrix} \mathbf{B}'^{21} \\ \mathbf{B}'^{22} \end{bmatrix} \mathbf{E}_{ij} + \mathbf{E}_{ji}[\mathbf{B}^{21} \;\; \mathbf{B}^{22}] \end{bmatrix} \quad \begin{pmatrix} i = 1, \ldots, q-p;\ j = 1, \ldots, q; \\ b^{p+i,j} \text{is a nonfixed parameter} \end{pmatrix}, \tag{4.6}$$

where $a^{ij} = (\mathbf{A}^{-1})_{ij}$ and $\mathbf{E}_{ij}$ is the matrix of an appropriate size whose (i, j) th element is 1 with other elements being 0.

The nonzero second partial derivatives of $\mathbf{\Sigma}$ with respect to $\boldsymbol{\theta}$ in notation are

$$\frac{\partial^2 \mathbf{\Sigma}}{\partial a^{ij}\, \partial a^{ik}} = \begin{bmatrix} \mathbf{E}_{jk} + \mathbf{E}_{kj} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix} \quad (i, j, k = 1, \ldots, p),$$

$$\frac{\partial^2 \mathbf{\Sigma}}{\partial b^{ij}\, \partial a^{ik}} = \begin{bmatrix} \mathbf{O} & \lambda_i \mathbf{E}_{kj} \\ \lambda_i \mathbf{E}_{jk} & \mathbf{O} \end{bmatrix} \quad (i, k = 1, \ldots, p;\ \ j = 1, \ldots, q),$$

$$\frac{\partial^2 \mathbf{\Sigma}}{\partial b^{ij}\, \partial b^{ik}} = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{E}_{jk} + \mathbf{E}_{kj} \end{bmatrix} \quad (i, j, k = 1, \ldots, q),$$

$$\frac{\partial^2 \mathbf{\Sigma}}{\partial \lambda_i\, \partial a^{ij}} = \begin{bmatrix} \mathbf{O} & \mathbf{E}_{ji}[\mathbf{B}^{11}\mathbf{B}^{12}] \\ \begin{bmatrix} \mathbf{B}'^{11} \\ \mathbf{B}'^{12} \end{bmatrix} \mathbf{E}_{ij} & \mathbf{O} \end{bmatrix} \quad (i, j = 1, \ldots, p),$$

$$\frac{\partial^2 \mathbf{\Sigma}}{\partial \lambda_i\, \partial b^{ij}} = \begin{bmatrix} \mathbf{O} & \mathbf{A}'^{-1}\mathbf{E}_{ij} \\ \mathbf{E}_{ji}\mathbf{A}^{-1} & \mathbf{O} \end{bmatrix} \quad (i = 1, \ldots, p;\ \ j = 1, \ldots, q), \tag{4.7}$$

with b^{ij} and b^{ik} being nonfixed parameters.

The nonzero third partial derivatives of $\mathbf{\Sigma}$ with respect to $\boldsymbol{\theta}$ in notation are

$$\frac{\partial^2 \mathbf{\Sigma}}{\partial \lambda_i\, \partial b^{ij}\, \partial a^{ik}} = \begin{bmatrix} \mathbf{O} & \mathbf{E}_{kj} \\ \mathbf{E}_{jk} & \mathbf{O} \end{bmatrix} \quad (i, k = 1, \ldots, p;\ \ j = 1, \ldots, q). \tag{4.8}$$

5. The transformed parameter estimators

As stated earlier, the parameters $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ were employed partially due to the tractability in the covariance structure. The transformed parameters $\mathbf{A}$ and $\mathbf{B}_1$ are also of interest as the coefficients of canonical variables. The asymptotic expansions of the distributions of $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}_1$ are available when the partial derivatives with respect to $\mathbf{s}$ are available as in $\hat{\boldsymbol{\theta}}$, which are given recursively with the results in the previous section as

$$\frac{\partial \hat{a}_{ij}}{\partial s_{cd}} = -\sum_{k,l=1}^{p} \hat{a}_{ik}\hat{a}_{lj}\frac{\partial \hat{a}^{kl}}{\partial s_{cd}}, \tag{5.1a}$$

$$\frac{\partial^2 \hat{a}_{ij}}{\partial s_{cd}\,\partial s_{ef}} = -\sum_{k,l=1}^{p} \left\{ \left(\frac{\partial \hat{a}_{ik}}{\partial s_{ef}}\hat{a}_{lj} + \hat{a}_{ik}\frac{\partial \hat{a}_{lj}}{\partial s_{ef}} \right) \frac{\partial \hat{a}^{kl}}{\partial s_{cd}} + \hat{a}_{ik}\hat{a}_{lj}\frac{\partial^2 \hat{a}^{kl}}{\partial s_{cd}\,\partial s_{ef}} \right\}, \tag{5.1b}$$

$$\begin{aligned}\frac{\partial^3 \hat{a}_{ij}}{\partial s_{cd}\,\partial s_{ef}\,\partial s_{gh}} = -\sum_{k,l=1}^{p} \Bigg\{ \Bigg(&\frac{\partial^2 \hat{a}_{ik}}{\partial s_{ef}\,\partial s_{gh}}\hat{a}_{lj} + \frac{\partial \hat{a}_{ik}}{\partial s_{ef}}\frac{\partial \hat{a}_{lj}}{\partial s_{gh}} + \frac{\partial \hat{a}_{ik}}{\partial s_{gh}}\frac{\partial \hat{a}_{lj}}{\partial s_{ef}} \\ &+ \hat{a}_{ik}\frac{\partial^2 \hat{a}_{lj}}{\partial s_{ef}\,\partial s_{gh}} \Bigg) \frac{\partial \hat{a}^{kl}}{\partial s_{cd}} + \left(\frac{\partial \hat{a}_{ik}}{\partial s_{ef}}\hat{a}_{lj} + \hat{a}_{ik}\frac{\partial \hat{a}_{lj}}{\partial s_{ef}} \right) \frac{\partial^2 \hat{a}^{kl}}{\partial s_{cd}\,\partial s_{gh}} \\ &+ \left(\frac{\partial \hat{a}_{ik}}{\partial s_{gh}}\hat{a}_{lj} + \hat{a}_{ik}\frac{\partial \hat{a}_{lj}}{\partial s_{gh}} \right) \frac{\partial^2 \hat{a}^{kl}}{\partial s_{cd}\,\partial s_{ef}} + \hat{a}_{ik}\hat{a}_{lj}\frac{\partial^3 \hat{a}^{kl}}{\partial s_{cd}\,\partial s_{ef}\,\partial s_{gh}} \Bigg\} \\ &(i, j = 1, \ldots, p;\ p+q \geqslant c \geqslant d \geqslant 1;\ p+q \\ &\geqslant e \geqslant f \geqslant 1;\ p+q \geqslant g \geqslant h \geqslant 1).\end{aligned} \tag{5.1c}$$

The partial derivatives of $\hat{\mathbf{B}}_1$ are similarly obtained as above. However, we should note that for the result of $\hat{\mathbf{B}}_1$ corresponding to (5.1b), the first partial derivatives of $\hat{\mathbf{B}}_2$ are also required and that for the result corresponding to (5.1c) the first and second partial derivatives of $\hat{\mathbf{B}}_2$ are required. Further, note that $\mathbf{B}$ is seen as a function of only nonfixed parameters in $\mathbf{B}^{-1}$.

Next, we deal with the summary statistics as transformed parameter estimators. The first one is the $-(2/n)$ log likelihood ratio statistic under fixed alternatives:

$$\hat{l}^* = -\sum_{e=1}^{p} \log(1-\hat{\lambda}_e^2). \tag{5.2}$$

For (5.2), we derive the partial derivatives of $\hat{\lambda}_e^2$ with respect to $\mathbf{s}$ using the results in the previous section as

$$\frac{\partial \hat{\lambda}_i^2}{\partial s_{kl}} = 2\hat{\lambda}_i\frac{\partial \hat{\lambda}_i}{\partial s_{kl}}, \quad \frac{\partial^2 \hat{\lambda}_i^2}{\partial s_{kl}\,\partial s_{cd}} = 2\frac{\partial \hat{\lambda}_i}{\partial s_{kl}}\frac{\partial \hat{\lambda}_i}{\partial s_{cd}} + 2\hat{\lambda}_i\frac{\partial^2 \hat{\lambda}_i}{\partial s_{kl}\,\partial s_{cd}},$$

$$\begin{aligned}\frac{\partial^3 \hat{\lambda}_i^2}{\partial s_{kl}\,\partial s_{cd}\,\partial s_{ef}} = 2\left(\frac{\partial \hat{\lambda}_i}{\partial s_{kl}}\frac{\partial^2 \hat{\lambda}_i}{\partial s_{cd}\,\partial s_{ef}} + \frac{\partial \hat{\lambda}_i}{\partial s_{cd}}\frac{\partial^2 \hat{\lambda}_i}{\partial s_{kl}\,\partial s_{ef}} + \frac{\partial \hat{\lambda}_i}{\partial s_{ef}}\frac{\partial^2 \hat{\lambda}_i}{\partial s_{cd}\,\partial s_{kl}} + \hat{\lambda}_i\frac{\partial^3 \hat{\lambda}_i}{\partial s_{kl}\,\partial s_{cd}\,\partial s_{ef}} \right) \\ (i = 1, \ldots, p;\ p+q \geqslant k \geqslant l \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1;\ p+q \geqslant e \geqslant f \geqslant 1).\end{aligned} \tag{5.3}$$

The partial derivatives of $\hat{l}^*$ with respect to $\mathbf{s}$ are given from (5.2) with (5.3) as

$$\frac{\partial \hat{l}^*}{\partial s_{ij}} = \sum_{e=1}^{p} \frac{1}{1-\hat{\lambda}_e^2} \frac{\partial \hat{\lambda}_e^2}{\partial s_{ij}}, \quad \frac{\partial^2 \hat{l}^*}{\partial s_{ij}\, \partial s_{kl}} = \sum_{e=1}^{p} \left\{ \frac{1}{(1-\hat{\lambda}_e^2)^2} \frac{\partial \hat{\lambda}_e^2}{\partial s_{ij}} \frac{\partial \hat{\lambda}_e^2}{\partial s_{kl}} + \frac{1}{1-\hat{\lambda}_e^2} \frac{\partial^2 \hat{\lambda}_e^2}{\partial s_{ij}\, \partial s_{kl}} \right\},$$

$$\begin{aligned} \frac{\partial^3 \hat{l}^*}{\partial s_{ij}\, \partial s_{kl}\, \partial s_{cd}} &= \sum_{e=1}^{p} \left\{ \frac{2}{(1-\hat{\lambda}_e^2)^3} \frac{\partial \hat{\lambda}_e^2}{\partial s_{ij}} \frac{\partial \hat{\lambda}_e^2}{\partial s_{kl}} \frac{\partial \hat{\lambda}_e^2}{\partial s_{cd}} + \frac{1}{(1-\hat{\lambda}_e^2)^2} \left(\frac{\partial \hat{\lambda}_e^2}{\partial s_{ij}} \frac{\partial^2 \hat{\lambda}_e^2}{\partial s_{kl}\, \partial s_{cd}} \right.\right. \\ &\quad \left.\left. + \frac{\partial \hat{\lambda}_e^2}{\partial s_{kl}} \frac{\partial^2 \hat{\lambda}_e^2}{\partial s_{ij}\, \partial s_{cd}} + \frac{\partial \hat{\lambda}_e^2}{\partial s_{cd}} \frac{\partial^2 \hat{\lambda}_e^2}{\partial s_{ij}\, \partial s_{kl}} \right) + \frac{1}{1-\hat{\lambda}_e^2} \frac{\partial^3 \hat{\lambda}_e^2}{\partial s_{ij}\, \partial s_{kl}\, \partial s_{cd}} \right\} \\ &\quad (p+q \geqslant i \geqslant j \geqslant 1;\ p+q \geqslant k \geqslant l \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1). \end{aligned} \tag{5.4}$$

The second summary statistic is Rozeboom's [28] between-set correlation coefficient or a matrix correlation between $\mathbf{x}$ and $\mathbf{y}$, whose squared sample value is defined as

$$\hat{\rho}_{XY}^2 = 1 - \frac{|\mathbf{S}|}{|\mathbf{S}_{XX}||\mathbf{S}_{YY}|} = 1 - \prod_{e=1}^{p} (1-\hat{\lambda}_e^2) = 1 - \exp(-\hat{l}^*) \tag{5.5}$$

(see also [18] for earlier works and statistical treatment of $\hat{\rho}_{XY}$ and $\sqrt{1-\hat{\rho}_{XY}^2}$ under normality). The partial derivatives of $\hat{\rho}_{XY}^2$ with respect to $\mathbf{s}$ are given from (5.5) with (5.4) as

$$\frac{\partial \hat{\rho}_{XY}^2}{\partial s_{ij}} = \exp(-\hat{l}^*) \frac{\partial \hat{l}^*}{\partial s_{ij}}, \quad \frac{\partial^2 \hat{\rho}_{XY}^2}{\partial s_{ij}\, \partial s_{kl}} = \exp(-\hat{l}^*) \left(-\frac{\partial \hat{l}^*}{\partial s_{ij}} \frac{\partial \hat{l}^*}{\partial s_{kl}} + \frac{\partial^2 \hat{l}^*}{\partial s_{ij}\, \partial s_{kl}} \right),$$

$$\begin{aligned} \frac{\partial^3 \hat{\rho}_{XY}^2}{\partial s_{ij}\, \partial s_{kl}\, \partial s_{cd}} &= \exp(-\hat{l}^*) \left(\frac{\partial \hat{l}^*}{\partial s_{ij}} \frac{\partial \hat{l}^*}{\partial s_{kl}} \frac{\partial \hat{l}^*}{\partial s_{cd}} - \frac{\partial \hat{l}^*}{\partial s_{ij}} \frac{\partial^2 \hat{l}^*}{\partial s_{kl}\, \partial s_{cd}} - \frac{\partial \hat{l}^*}{\partial s_{kl}} \frac{\partial^2 \hat{l}^*}{\partial s_{ij}\, \partial s_{cd}} \right. \\ &\quad \left. - \frac{\partial \hat{l}^*}{\partial s_{cd}} \frac{\partial^2 \hat{l}^*}{\partial s_{ij} \partial s_{kl}} + \frac{\partial^3 \hat{l}^*}{\partial s_{ij} \partial s_{kl}\, \partial s_{cd}} \right) \\ &\quad (p+q \geqslant i \geqslant j \geqslant 1;\ p+q \geqslant k \geqslant l \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1). \end{aligned} \tag{5.6}$$

Finally, the partial derivatives of nonsquared Rozeboom's coefficient are given from (5.6) as

$$\frac{\partial \hat{\rho}_{XY}}{\partial s_{ij}} = \frac{\hat{\rho}_{XY}^{-1}}{2} \frac{\partial \hat{\rho}_{XY}^2}{\partial s_{ij}}, \quad \frac{\partial^2 \hat{\rho}_{XY}}{\partial s_{ij}\, \partial s_{kl}} = -\frac{\hat{\rho}_{XY}^{-3}}{4} \frac{\partial \hat{\rho}_{XY}^2}{\partial s_{ij}} \frac{\partial \hat{\rho}_{XY}^2}{\partial s_{kl}} + \frac{\hat{\rho}_{XY}^{-1}}{2} \frac{\partial^2 \hat{\rho}_{XY}^2}{\partial s_{ij}\, \partial s_{kl}},$$

$$\begin{aligned} \frac{\partial^3 \hat{\rho}_{XY}}{\partial s_{ij}\, \partial s_{kl}\, \partial s_{cd}} &= \frac{3}{8} \hat{\rho}_{XY}^{-5} \frac{\partial \hat{\rho}_{XY}^2}{\partial s_{ij}} \frac{\partial \hat{\rho}_{XY}^2}{\partial s_{kl}} \frac{\partial \hat{\rho}_{XY}^2}{\partial s_{cd}} - \frac{\hat{\rho}_{XY}^{-3}}{4} \left(\frac{\partial \hat{\rho}_{XY}^2}{\partial s_{ij}} \frac{\partial^2 \hat{\rho}_{XY}^2}{\partial s_{kl}\, \partial s_{cd}} \right. \\ &\quad \left. + \frac{\partial \hat{\rho}_{XY}^2}{\partial s_{kl}} \frac{\partial^2 \hat{\rho}_{XY}^2}{\partial s_{ij}\, \partial s_{cd}} + \frac{\partial \hat{\rho}_{XY}^2}{\partial s_{cd}} \frac{\partial^2 \hat{\rho}_{XY}^2}{\partial s_{ij}\, \partial s_{kl}} \right) + \frac{\hat{\rho}_{XY}^{-1}}{2} \frac{\partial^2 \hat{\rho}_{XY}^2}{\partial s_{ij} \partial s_{kl}\, \partial s_{cd}} \\ &\quad (p+q \geqslant i \geqslant j \geqslant 1;\ p+q \geqslant k \geqslant l \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1). \end{aligned} \tag{5.7}$$

6. Numerical examples

In this section, two numerical examples are illustrated. First, a real data numerical example is shown with simulations using Lawley and Maxwell's [22, p. 66] correlation matrix for six school subjects ($N = 220$). The correlation matrix was regarded as a population covariance matrix, where the six variables are reordered such that X_1, X_2 and $Y_1, \ldots, Y_4$ stand for the scores of English, Arithmetic, Gaelic, History, Algebra and Geometry, respectively. In the initial set of parameters $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$ and $\boldsymbol{\lambda}$, the lower-left element of $[\mathbf{B}^{21}\ \mathbf{B}^{22}]$ was set to 0 in order to remove the rotational indeterminacy.

Simulations were performed to have the true cumulants of the non-Studentized and NT Studentized estimators under normality and nonnormality. (The case of the ADF Studentized estimator will be dealt with later in the second numerical example.) The minor indeterminacy of the signs of the rows of $[\hat{\mathbf{A}}^{-1}\ \hat{\mathbf{B}}^{11}\ \hat{\mathbf{B}}^{12}]$ and $[\hat{\mathbf{B}}^{21}\ \hat{\mathbf{B}}^{22}]$ in simulations was removed to search the patterns most similar to the population ones. Random observations were generated by the relationship $(\mathbf{x}', \mathbf{y}')' = \boldsymbol{\Sigma}_1\mathbf{z}$ with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_1'$, where $\boldsymbol{\Sigma}_1$ is given from the Cholesky decomposition of $\boldsymbol{\Sigma}$ and the elements of the random vector $\mathbf{z}$ have independent distributions with unit variances. The chi-square distribution with 1 degree of freedom followed by standardization with unit variance was used for nonnormal data. A set of parameter estimates was obtained from the sample covariance matrix based on the random observations with the same sample size as the real one. This was replicated 1,000,000 times, which gave 1,000,000 estimates for each (transformed) parameter.

Tables 1 and 2 show the results. Table 1 and part of Table 2 contain the result of the non-Studentized parameters while the remaining part of Table 2 gives the result of the NT Studentized parameters. In the tables, simulated cumulants are shown, which were given by the k-statistics (unbiased estimators of cumulants) multiplied by the appropriate powers of n for ease of comparison with the corresponding theoretical values. In Table 2, HSE denotes the higher-order

Table 1
Theoretical and simulated cumulants of the non-Studentized estimators in six school subjects ($N = 220$)

Parameter	$\alpha_2^{1/2}$ (dispersion)				α_1 (bias)			
	Normal		Chi-square ($df = 1$)		Normal		Chi-square ($df = 1$)	
	Th.	Sim.	Th.	Sim.	Th.	Sim.	Th.	Sim.
$\lambda_1 = .68$	.54	.53	.80	.77	1.56	1.58	2.92	2.79
$\lambda_2 = .34$	.88	.86	1.01	.97	1.95	1.96	1.53	1.65
$a_{11} = .42$	1.48	1.53	1.84	1.87	−.22	−.24	1.09	1.01
$b_{11} = .26$	1.38	1.41	2.21	2.16	−.45	−.46	1.07	.88
$l^* = .74$	1.52	1.55	2.39	2.43	8.00	8.13	13.22	13.11
$\rho_{XY} = .72$	.50	.49	.79	.74	2.08	2.07	2.98	2.88
	α_3 (skewness)				α_4 (kurtosis)			
λ_1	−.6	−.6	−2.2	−2.0	2	2	10	8
λ_2	−1.4	−1.2	−1.8	−1.1	1	−1	11	−2
a_{11}	−2.6	−4.1	13.2	15.8	145	244	416	591
b_{11}	−.9	−1.0	61.2	52.4	74	97	711	615
l^*	8.4	8.7	79.9	78.5	34	34	1834	1703
ρ_{XY}	−.5	−.5	−2.3	−1.9	2	2	10	8

Note: Th., Theoretical values; Sim., simulated values.

Table 2
Theoretical and simulated HSEs of the non-Studentized estimators and cumulants of the NT Studentized ones in six school subjects ($N = 220$)

Parameter	HSE/SE	SD/SE	HSE/SE	SD/SE	$\alpha_{NT2}'^{1/2}$		$\alpha_{NT2}''^{1/2}$(dispersion)	
	Normal		Chi-square ($df = 1$)		Normal		Chi-square ($df = 1$)	
	Th.	Sim.	Th.	Sim.	Th.	Sim.	Th.	Sim.
λ_1	.982	.981	.951	.956	1	1.035	1.490	1.606
λ_2	.976	.975	.954	.958	1	1.003	1.142	1.132
a_{11}	1.033	1.037	1.010	1.017	1	1.030	1.249	1.297
b_{11}	1.020	1.021	.968	.979	1	1.023	1.597	1.626
l^*	1.017	1.017	1.015	1.017	1	.991	1.574	1.537
ρ_{XY}	.973	.973	.932	.941	1	1.046	1.574	1.707
	α_{NT1}'		α_{NT1}'' (bias)		α_{NT3}'		α_{NT3}''(skewness)	
λ_1	4.24	4.39	8.42	8.59	4.1	4.9	26.0	36.3
λ_2	2.89	2.94	2.63	2.76	2.1	2.3	4.3	5.3
a_{11}	.15	.17	.33	.31	.9	1.1	.3	1.1
b_{11}	−.21	−.21	−.71	−.70	.4	.4	.4	1.8
l^*	4.46	4.50	6.72	6.65	−2.4	−2.2	−6.7	−5.7
ρ_{XY}	5.57	5.75	9.46	9.75	4.2	5.2	34.1	48.5

Note: Th., Theoretical values; Sim., simulated values; HSE $= \sqrt{(\alpha_2/n) + (\Delta\alpha_2/n^2)}$; SE $= \sqrt{\alpha_2/n}$; SD, standard deviation from simulation.

asymptotic standard error of a non-Studentized estimator or $\sqrt{(\alpha_2/n) + (\Delta\alpha_2/n^2)}$, while SD is the corresponding simulated value or the standard deviation from the simulation, and SE is the usual asymptotic standard error or $\sqrt{\alpha_2/n}$ for the non-Studentized estimator.

The results are shown for selected (transformed) parameters. In the tables, only λ_1 and λ_2 are initial parameters while other parameters are transformed ones. The results of the tables show that the theoretical or asymptotic cumulants are similar to their corresponding simulated values and that the absolute values of the cumulants under nonnormality are mostly substantially larger than the corresponding values under normality. This stems mainly from the large kurtosis of the chi-square distribution. It is of interest to find in Table 2 that some of the ratios HSE/ASE and SD/SE are different from 1 by 5–6% under nonnormality while the differences are relatively small under normality.

Table 3 shows the result of the simulation for confidence intervals based on sample cumulants of NT Studentized estimators under normality. The simulation was performed as in Tables 1 and 2 with the reduced number of replications being 100,000 due to the excessive computation time required to have sample cumulants in each replication. The confidence intervals with various nominal confidence coefficients were constructed in three ways: the usual normal approximation, the Cornish–Fisher expansion and Hall's method by variable transformation (see (3.12) and (3.13)). The values in the table show the proportions of the population parameters below the lower limits of confidence intervals in the simulation. The table shows that except for a_{11} and b_{11}, the simulated proportions by the normal approximation are not satisfactory while the corresponding proportions by the other two methods improve considerably over the results by the normal approximation. The results of the two improved methods are similar.

Table 3
Simulated proportions of true parameters below the lower limits of the confidence intervals based on the NT Studentized estimators under normality in six school subjects ($N = 220$)

Parameter	Method	Nominal values						
		.0050	.0250	.1000	.5000	.9000	.9750	.9950
λ_1	N*	.0212	.0615	.1657	.5961	.9462	.9919	.9994
	C–F	.0068	.0306	.1084	.5016	.8930	.9700	.9934
	Hall	.0057	.0290	.1080	.5015	.8932	.9710	.9941
λ_2	N*	.0119	.0441	.1399	.5690	.9350	.9878	.9983
	C–F	.0059	.0287	.1082	.5020	.8929	.9708	.9934
	Hall	.0056	.0281	.1080	.5020	.8929	.9710	.9935
a_{11}	N*	.0081	.0314	.1080	.5008	.8992	.9724	.9936
	C–F	.0065	.0280	.1044	.5013	.8973	.9715	.9933
	Hall	.0063	.0278	.1043	.5013	.8973	.9715	.9933
b_{11}	N*	.0068	.0277	.1017	.4926	.8946	.9712	.9936
	C–F	.0066	.0274	.1029	.4952	.8942	.9709	.9936
	Hall	.0066	.0273	.1029	.4952	.8942	.9709	.9936
l^*	N*	.0081	.0428	.1596	.6303	.9422	.9860	.9967
	C–F	.0048	.0252	.0996	.5020	.9017	.9765	.9953
	Hall	.0047	.0250	.0995	.5020	.9018	.9771	.9957
ρ_{XY}	N*	.0256	.0722	.1893	.6303	.9560	.9934	.9993
	C–F	.0071	.0314	.1119	.5050	.8912	.9697	.9926
	Hall	.0058	.0296	.1115	.5049	.8915	.9709	.9934

Note: N*, normal approximation; C–F, Cornish–Fisher expansion; Hall, Hall's method by variable transformation.

Table 4 shows the overall errors of the asymptotic distribution functions of the estimators standardized by the population ADF standard errors. The true values were defined by the simulated distribution functions in the data used in Table 1 at the 40 points, $-3.8, -3.6, \ldots, 4.0$. The theoretical or asymptotic values were given in four ways: the normal approximation, the single-term Edgeworth expansion up to order $O(n^{-1/2})$, the two-term Edgeworth expansion up to order $O(n^{-1})$, and Hall's method by variable transformation. The overall error was defined as the square root of the mean of the squared differences of the true (simulated) values and the corresponding theoretical values of a distribution function over the 40 points. In the table we find that the two-term Edgeworth expansions have the smallest overall errors while the relative error sizes of the single-term Edgeworth expansion and Hall's method depend on parameters.

Fig. 1 illustrates the errors of the distribution functions of the ADF standardized (standardized using the ADF standard error) estimators by the four methods at the 40 points whose summary results were given in Table 4. We find that the two-term Edgeworth expansion gives very small errors for λ_1 and ρ_{XY} in all area.

In the second numerical example, the ADF Studentized estimator (see (3.8)) is dealt with. The sample ADF standard error is given by $n^{-1/2}\{(\partial\hat{\theta}/\partial\mathbf{s}')\hat{\mathbf{\Omega}}(\partial\hat{\theta}/\partial\mathbf{s})\}^{1/2}$. For $\mathbf{\Omega}$, the unbiased estimator is available [8,20]. However, the following consistent estimator is simple and asymptotically

Table 4
$10^5 \times$ root mean square errors of the asymptotic distribution functions of the ADF standardized estimators in six school subjects ($N = 220$)

Data	Method	Parameters					
		λ_1	λ_2	a_{11}	b_{11}	l^*	ρ_{XY}
Normal	N*	4186	3062	391	442	6510	5759
	E1	202	127	352	212	892	444
	E2	58	32	38	19	118	35
	Hall	315	317	352	210	242	431
Chi-square ($df = 1$)	N*	4956	2296	567	886	6403	5263
	E1	283	447	156	354	706	345
	E2	178	153	108	170	159	188
	Hall	684	557	140	359	191	876

Note: N*, normal approximation; E1, the single-term Edgeworth expansion; E2, the two-term Edgeworth expansion; Hall, Hall's method by variable transformation.

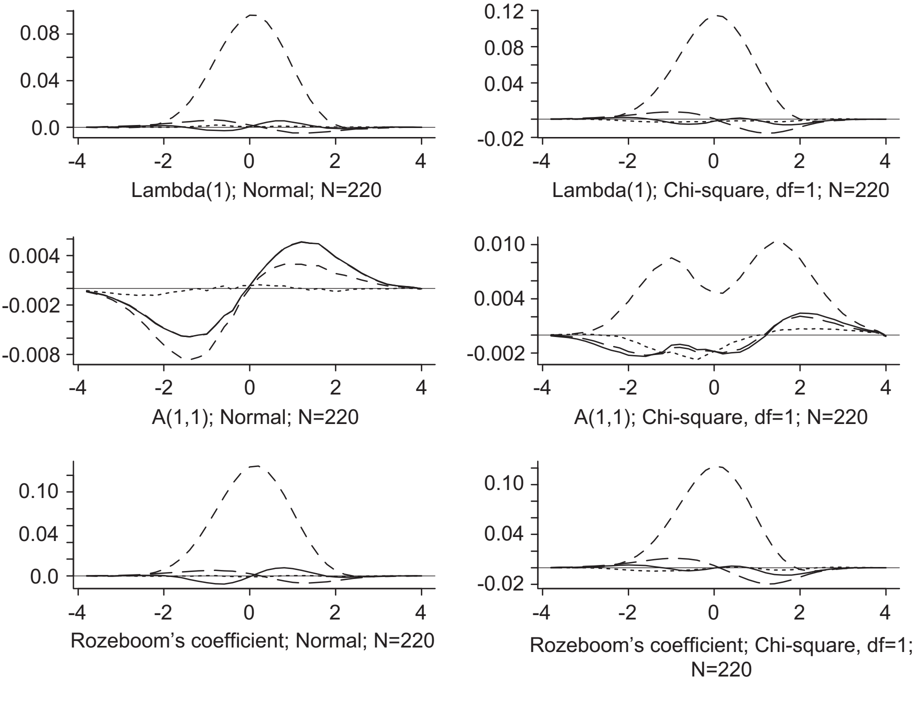

Fig. 1. Errors of the distribution functions of the ADF standardized estimators in six school subjects (dashed lines are the standard normal distribution; solid lines, the single term Edgeworth expansion; dotted lines, the two-term Edgeworth expansion; long dashed lines, Hall's variable transformation).

Table 5
Theoretical and simulated cumulants of the ADF Studentized estimators in four-variable artificial data ($N = 300$)

Parameter	Population value		$(\alpha_2^{1/2}/\alpha_{\mathrm{NT2}}^{1/2})$		$\alpha_2'^{1/2}$ (dispersion)			
	Normal		Chi-square ($df = 3$)		Normal		Chi-square ($df = 3$)	
	Th.	Sim.	Th.	Sim.	Th.	Sim.	Th.	Sim.
λ_1	.8		(1.30)		1	1.034	1	1.080
λ_2	.4		(1.08)		1	1.024	1	1.044
a_{11}	.707		(1.24)		1	1.019	1	1.046
b_{11}	.707		(1.33)		1	1.019	1	1.058
l^*	1.196		(1.35)		1	1.014	1	1.048
ρ_{XY}	.835		(1.35)		1	1.036	1	1.082
	α_1' (bias)				α_3' (skewness)			
λ_1	2.60	2.69	3.20	3.34	4.8	5.5	6.4	8.2
λ_2	.30	.30	.08	.06	2.4	2.6	2.7	3.0
a_{11}	−.07	−.08	.26	.16	−.3	−.3	.5	−.2
b_{11}	−.07	−.09	.45	.33	−.3	−.3	.5	−.3
l^*	1.73	1.77	1.60	1.66	−1.5	−1.6	−2.4	−2.7
ρ_{XY}	2.81	2.91	3.07	3.24	5.0	5.8	6.5	8.5

Note: Th., Theoretical values, Sim., Simulated values.

equivalent to the unbiased one, and is used in this paper.

$$(\hat{\mathbf{\Omega}})_{ab,cd} = \frac{1}{N}\sum_{i=1}^{N}(X_{ia}-\overline{X}_a)(X_{ib}-\overline{X}_b)(X_{ic}-\overline{X}_c)(X_{id}-\overline{X}_d) - s_{ab}s_{cd}$$
$$(p+q \geqslant a \geqslant b \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1), \tag{6.1}$$

where X_{ia} is the ith observation of variable X_a and $\overline{X}_a = (1/N)\sum_{i=1}^{N} X_{ia}$.

Since (6.1) uses the sample's fourth moment which tends to be unstable with small to moderate sample sizes, a tractable artificial data set with $p = 2$, $q = 2$ and

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \text{symmetric} & & \\ 0 & 1 & & \\ 0.6 & 0.2 & 1 & \\ 0.2 & 0.6 & 0 & 1 \end{bmatrix} \tag{6.2}$$

is used. Simulations were performed as in the first example. Nonnormal observations were similarly generated though the chi-square distribution with 3 degrees of freedom is used for each element of $\mathbf{z}$ (recall $(\mathbf{x}', \mathbf{y}')' = \mathbf{\Sigma}_1\mathbf{z}$ with $\mathbf{\Sigma} = \mathbf{\Sigma}_1\mathbf{\Sigma}_1'$) to have moderately nonnormal data for stability. Though X_1 and X_2 (Y_1 and Y_2) can be exchanged without changing $\mathbf{\Sigma}$, they have different distributions under nonnormality due to $\mathbf{\Sigma} = \mathbf{\Sigma}_1\mathbf{\Sigma}_1'$.

Table 5 shows the population and simulated asymptotic cumulants of the ADF Studentized estimators with $N = 300$, where 100,000 replications were used as before. Note that the results under normality are included for comparison though the simulated values were given using the ADF estimates instead of the NT ones. It is found that the theoretical values are reasonably close to their corresponding simulated ones. The actual standard deviations of the ADF Studentized estimators are somewhat higher than the unit asymptotic value especially in the nonnormal case,

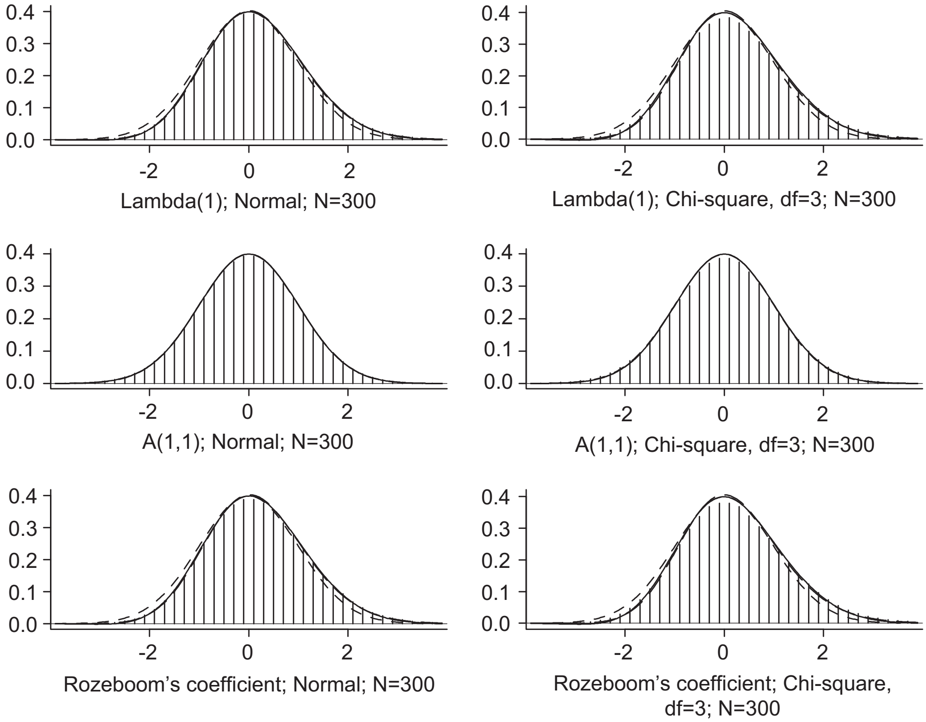

Fig. 2. Theoretical (curved lines) and simulated (histograms) distributions of the ADF Studentized estimators in four-variable artificial data (dashed lines are the standard normal distribution; solid lines, the single-term Edgeworth expansion; long dashed lines, Hall's variable transformation).

which indicates some room of improvement by the higher-order asymptotic standard error of the ADF Studentized estimator even when N is as large as 300. Fig. 2 illustrates the simulated and theoretical densities of the ADF Studentized estimators of λ_1, a_{11} and ρ_{XY} under normality and nonnormality. The figure shows the improvements given by the single-term Edgeworth expansion and Hall's method in the cases of λ_1 and ρ_{XY}.

For the simulation of confidence intervals corresponding to that in Table 3, the sample moments up to the sixth order are required for direct application of the Cornish–Fisher expansion and Hall's method (see (3.11) and (3.12)). This was tried but stable results were not available while the results by the usual normal approximation using the ADF standard errors were relatively stable. In Table 5 note the similarity of the cumulants of the ADF Studentized estimators under normality and nonnormality in spite of the substantial difference of $\alpha_2^{1/2}$ and $\alpha_{\mathrm{NT2}}^{1/2}$ for most of the parameter estimators (see the values of $\alpha_2^{1/2}/\alpha_{\mathrm{NT2}}^{1/2}$ shown parenthetically in Table 5). Considering the similarity with $\alpha_1' = \alpha_{\mathrm{NT1}}'$ and $\alpha_3' = \alpha_{\mathrm{NT3}}'$ under normality (which can be algebraically derived), we constructed the confidence intervals using $\hat{\alpha}_2^{1/2}$ with stable $\hat{\alpha}_{\mathrm{NT1}}'$ and $\hat{\alpha}_{\mathrm{NT3}}'$ in place of unstable $\hat{\alpha}_1'$ and $\hat{\alpha}_3'$. That is, for the confidence interval by the Cornish–Fisher expansion, we used

$$\hat{\theta} + [\pm z_{\tilde{\alpha}/2} - n^{-1/2}\{\hat{\alpha}_{\mathrm{NT1}}' + (\hat{\alpha}_{\mathrm{NT3}}'/6)(z_{\tilde{\alpha}/2}^2 - 1)\}]n^{-1/2}\hat{\alpha}_2^{1/2} \tag{6.3}$$

Table 6
Simulated proportions of true parameters below the lower limits of the confidence intervals based on the ADF Studentized estimators under nonnormality (chi-square, $df = 3$) in four-variable artificial data ($N = 300$)

Parameter	Method	Nominal values						
		.0050	.0250	.1000	.5000	.9000	.9750	.9950
λ_1	N*	.0223	.0600	.1545	.5517	.9247	.9826	.9981
	C–F	.0101	.0384	.1214	.5129	.8940	.9698	.9922
	Hall	.0087	.0368	.1209	.5129	.8945	.9709	.9931
λ_2	N*	.0101	.0353	.1124	.4913	.8953	.9741	.9944
	C–F	.0067	.0291	.1064	.4939	.8897	.9699	.9927
	Hall	.0064	.0286	.1063	.4939	.8897	.9701	.9927
a_{11}	N*	.0081	.0319	.1085	.5032	.8938	.9686	.9921
	C–F	.0079	.0320	.1092	.5046	.8945	.9690	.9923
	Hall	.0079	.0320	.1092	.5046	.8945	.9690	.9923
b_{11}	N*	.0085	.0329	.1125	.5070	.8927	.9681	.9909
	C–F	.0089	.0337	.1137	.5074	.8936	.9692	.9915
	Hall	.0089	.0336	.1137	.5074	.8937	.9692	.9915
l^*	N*	.0075	.0343	.1278	.5481	.9034	.9715	.9921
	C–F	.0068	.0298	.1103	.5049	.8888	.9679	.9918
	Hall	.0067	.0298	.1102	.5049	.8888	.9681	.9919
ρ_{XY}	N*	.0225	.0601	.1547	.5481	.9244	.9857	.9981
	C–F	.0101	.0373	.1208	.5054	.8891	.9687	.9922
	Hall	.0086	.0352	.1205	.5053	.8895	.9700	.9932

Note: N*, Normal approximation; C–F, Cornish–Fisher expansion; Hall, Hall's method by variable transformation. In the methods of C–F and Hall, $\hat{\alpha}'_{\mathrm{NT1}}$ and $\hat{\alpha}'_{\mathrm{NT3}}$ with $\hat{\alpha}'_2$ are used.

in place of (3.11). The similar replacement was also employed in (3.12) for Hall's method. Of course, this replacement does not guarantee the improvement of the accuracy order in the asymptotic confidence coefficient beyond that by the usual normal approximation using only $\hat{\alpha}_2^{1/2}$. However, substantial improvement was expected in finite samples.

Table 6 shows the results obtained by 100,000 confidence intervals for each parameter, where the proportions by the Cornish–Fisher expansion and Hall's method were given by the above replacement. As expected, improvement of accuracy over the normal approximation was obtained especially for the parameters that are functions of canonical correlations.

7. Other transformed estimators

The structure covariances, used as initial parameters, can also be defined in crossed cases i.e., the covariances ($\mathbf{B}_1'\mathbf{\Sigma}_{YX}$) between $\mathbf{g}$ and $\mathbf{x}'$, and those ($\mathbf{A}'\mathbf{\Sigma}_{XY}$) between $\mathbf{f}$ and $\mathbf{y}'$. It can be shown that $\mathbf{B}_1'\mathbf{\Sigma}_{YX} = \mathbf{\Lambda A}'\mathbf{\Sigma}_{XX} = \mathbf{\Lambda A}^{-1}$ and $\mathbf{A}'\mathbf{\Sigma}_{XY} = \mathbf{\Lambda B}_1'\mathbf{\Sigma}_{YY} = \mathbf{\Lambda}[\mathbf{B}^{11}\mathbf{B}^{12}]$. Since the partial derivatives of $\hat{\mathbf{A}}'^{-1}$ and $[\hat{\mathbf{B}}^{11}\ \hat{\mathbf{B}}^{12}]$ are available, it is convenient to use these relationships. Let

$$\mathrm{Cov}\{(\mathbf{f}', \mathbf{g}')', (\mathbf{x}', \mathbf{y}')\} \equiv \mathbf{U} = \begin{bmatrix} \mathbf{U}_{fX} & \mathbf{U}_{fY} \\ \mathbf{U}_{gX} & \mathbf{U}_{gY} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{\Lambda}[\mathbf{B}^{11}\mathbf{B}^{12}] \\ \mathbf{\Lambda A}^{-1} & [\mathbf{B}^{11}\,\mathbf{B}^{12}] \end{bmatrix}. \tag{7.1}$$

Then, we have the partial derivatives of $\hat{\mathbf{U}}_{gX}$:

$$\frac{\partial \hat{\mathbf{U}}_{gX}}{\partial s_{ij}} = \frac{\partial \hat{\mathbf{\Lambda}}}{\partial s_{ij}} \hat{\mathbf{A}}^{-1} + \hat{\mathbf{\Lambda}} \frac{\partial \hat{\mathbf{A}}^{-1}}{\partial s_{ij}},$$

$$\frac{\partial^2 \hat{\mathbf{U}}_{gX}}{\partial s_{ij}\, \partial s_{kl}} = \frac{\partial^2 \hat{\mathbf{\Lambda}}}{\partial s_{ij}\, \partial s_{kl}} \hat{\mathbf{A}}^{-1} + \frac{\partial \hat{\mathbf{\Lambda}}}{\partial s_{ij}} \frac{\partial \hat{\mathbf{A}}^{-1}}{\partial s_{kl}} + \frac{\partial \hat{\mathbf{\Lambda}}}{\partial s_{kl}} \frac{\partial \hat{\mathbf{A}}^{-1}}{\partial s_{ij}} + \hat{\mathbf{\Lambda}} \frac{\partial^2 \hat{\mathbf{A}}^{-1}}{\partial s_{ij}\, \partial s_{kl}},$$

$$\begin{aligned}\frac{\partial^3 \hat{\mathbf{U}}_{gX}}{\partial s_{ij}\, \partial s_{kl}\, \partial s_{cd}} = {} & \frac{\partial^3 \hat{\mathbf{\Lambda}}}{\partial s_{ij}\, \partial s_{kl}\, \partial s_{cd}} \hat{\mathbf{A}}^{-1} + \frac{\partial^2 \hat{\mathbf{\Lambda}}}{\partial s_{ij} \partial s_{kl}} \frac{\partial \hat{\mathbf{A}}^{-1}}{\partial s_{cd}} + \frac{\partial^2 \hat{\mathbf{\Lambda}}}{\partial s_{ij}\, \partial s_{cd}} \frac{\partial \hat{\mathbf{A}}^{-1}}{\partial s_{kl}} + \frac{\partial^2 \hat{\mathbf{\Lambda}}}{\partial s_{kl}\, \partial s_{cd}} \frac{\partial \hat{\mathbf{A}}^{-1}}{\partial s_{ij}} \\ & + \frac{\partial \hat{\mathbf{\Lambda}}}{\partial s_{ij}} \frac{\partial^2 \hat{\mathbf{A}}^{-1}}{\partial s_{kl}\, \partial s_{cd}} + \frac{\partial \hat{\mathbf{\Lambda}}}{\partial s_{kl}} \frac{\partial^2 \hat{\mathbf{A}}^{-1}}{\partial s_{ij}\, \partial s_{cd}} + \frac{\partial \hat{\mathbf{\Lambda}}}{\partial s_{cd}} \frac{\partial^2 \hat{\mathbf{A}}^{-1}}{\partial s_{ij}\, \partial s_{kl}} + \hat{\mathbf{\Lambda}} \frac{\partial^3 \hat{\mathbf{A}}^{-1}}{\partial s_{ij} \partial s_{kl}\, \partial s_{cd}} \\ & (p+q \geqslant i \geqslant j \geqslant 1;\ p+q \geqslant k \geqslant l \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1). \end{aligned} \tag{7.2}$$

The results for $\hat{\mathbf{U}}_{fY}$ are similarly obtained.

The structures defined in covariances may also be defined in correlations including crossed ones as structure correlations. Let

$$\mathrm{Cor}\{(\mathbf{f}', \mathbf{g}')', (\mathbf{x}', \mathbf{y}')\} \equiv \mathbf{V} = \mathbf{U}\{\mathrm{Diag}(\mathbf{\Sigma})\}^{-1/2}, \tag{7.3}$$

where $\mathrm{Diag}(\cdot)$ denotes the diagonal matrix with the diagonal elements of an argument matrix. Then, the partial derivatives of $\hat{\mathbf{V}}$ are given using (7.2) as

$$\frac{\partial \hat{\mathbf{V}}}{\partial s_{ij}} = \frac{\partial \hat{\mathbf{U}}}{\partial s_{ij}} \{\mathrm{Diag}(\mathbf{S})\}^{-1/2} - \frac{1}{2} \hat{\mathbf{U}} \delta_{ij} \mathbf{E}_{ii} \{\mathrm{Diag}(\mathbf{S})\}^{-3/2},$$

$$\begin{aligned}\frac{\partial^2 \hat{\mathbf{V}}}{\partial s_{ij}\, \partial s_{kl}} = {} & \frac{\partial^2 \hat{\mathbf{U}}}{\partial s_{ij}\, \partial s_{kl}} \{\mathrm{Diag}(\mathbf{S})\}^{-1/2} - \frac{1}{2} \left(\frac{\partial \hat{\mathbf{U}}}{\partial s_{kl}} \delta_{ij} \mathbf{E}_{ii} + \frac{\partial \hat{\mathbf{U}}}{\partial s_{ij}} \delta_{kl} \mathbf{E}_{kk} \right) \{\mathrm{Diag}(\mathbf{S})\}^{-3/2} \\ & + \frac{3}{4} \hat{\mathbf{U}} \delta_{ij} \delta_{kl} \delta_{ik} \mathbf{E}_{ii} \{\mathrm{Diag}(\mathbf{S})\}^{-5/2}, \end{aligned}$$

$$\begin{aligned}\frac{\partial^3 \hat{\mathbf{V}}}{\partial s_{ij}\, \partial s_{kl}\, \partial s_{cd}} = {} & \frac{\partial^3 \hat{\mathbf{U}}}{\partial s_{ij} \partial s_{kl}\, \partial s_{cd}} \{\mathrm{Diag}(\mathbf{S})\}^{-1/2} - \frac{1}{2} \left(\frac{\partial^2 \hat{\mathbf{U}}}{\partial s_{kl}\, \partial s_{cd}} \delta_{ij} \mathbf{E}_{ii} \right. \\ & \left. + \frac{\partial^2 \hat{\mathbf{U}}}{\partial s_{ij}\, \partial s_{cd}} \delta_{kl} \mathbf{E}_{kk} + \frac{\partial^2 \hat{\mathbf{U}}}{\partial s_{ij}\, \partial s_{kl}} \delta_{cd} \mathbf{E}_{cc} \right) \{\mathrm{Diag}(\mathbf{S})\}^{-3/2} \\ & + \frac{3}{4} \left(\frac{\partial \hat{\mathbf{U}}}{\partial s_{kl}} \delta_{ij} \delta_{cd} \delta_{ic} \mathbf{E}_{ii} + \frac{\partial \hat{\mathbf{U}}}{\partial s_{cd}} \delta_{kl} \delta_{ij} \delta_{ik} \mathbf{E}_{kk} + \frac{\partial \hat{\mathbf{U}}}{\partial s_{ij}} \delta_{cd} \delta_{kl} \delta_{kc} \mathbf{E}_{cc} \right) \\ & \times \{\mathrm{Diag}(\mathbf{S})\}^{-5/2} - \frac{15}{8} \hat{\mathbf{U}} \delta_{ij} \delta_{kl} \delta_{ik} \delta_{cd} \delta_{ic} \mathbf{E}_{ii} \{\mathrm{Diag}(\mathbf{S})\}^{-7/2} \\ & (p+q \geqslant i \geqslant j \geqslant 1;\ p+q \geqslant k \geqslant l \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1), \end{aligned} \tag{7.4}$$

where δ_{ij} is the Kronecker delta.

The coefficients of canonical variables with respect to standardized observable variables are defined as

$$\mathbf{A}_\rho = \mathbf{D}_X\mathbf{A} \quad \text{and} \quad \mathbf{B}_{1\rho} = \mathbf{D}_Y\mathbf{B}_1, \tag{7.5}$$

with $\mathbf{D}_X \equiv \{\mathrm{Diag}(\mathbf{\Sigma}_{XX})\}^{1/2}$ and $\mathbf{D}_Y \equiv \{\mathrm{Diag}(\mathbf{\Sigma}_{YY})\}^{1/2}$. The partial derivatives of $\hat{\mathbf{A}}_\rho$ are given by using those of $\hat{\mathbf{A}}$ as

$$\begin{aligned}
\frac{\partial \hat{\mathbf{A}}_\rho}{\partial s_{ij}} &= \frac{\delta_{ij}}{2}\mathbf{E}_{ii}\hat{\mathbf{D}}_X^{-1}\hat{\mathbf{A}} + \hat{\mathbf{D}}_X\frac{\partial \hat{\mathbf{A}}}{\partial s_{ij}},\\
\frac{\partial^2 \hat{\mathbf{A}}_\rho}{\partial s_{ij}\,\partial s_{kl}} &= -\frac{1}{4}\delta_{ij}\delta_{kl}\delta_{ik}\mathbf{E}_{ii}\hat{\mathbf{D}}_X^{-3}\hat{\mathbf{A}} + \frac{\delta_{ij}}{2}\mathbf{E}_{ii}\hat{\mathbf{D}}_X^{-1}\frac{\partial \hat{\mathbf{A}}}{\partial s_{kl}} + \frac{\delta_{kl}}{2}\mathbf{E}_{kk}\hat{\mathbf{D}}_X^{-1}\frac{\partial \hat{\mathbf{A}}}{\partial s_{ij}} + \hat{\mathbf{D}}_X\frac{\partial^2 \hat{\mathbf{A}}}{\partial s_{ij}\,\partial s_{kl}},\\
\frac{\partial^3 \hat{\mathbf{A}}_\rho}{\partial s_{ij}\,\partial s_{kl}\,\partial s_{cd}} &= \frac{3}{8}\delta_{ij}\delta_{kl}\delta_{ik}\delta_{cd}\delta_{ic}\mathbf{E}_{ii}\hat{\mathbf{D}}_X^{-5}\hat{\mathbf{A}} - \frac{1}{4}\delta_{ij}\delta_{kl}\delta_{ik}\mathbf{E}_{ii}\hat{\mathbf{D}}_X^{-3}\frac{\partial \hat{\mathbf{A}}}{\partial s_{cd}}\\
&\quad -\frac{1}{4}\delta_{ij}\delta_{cd}\delta_{ic}\mathbf{E}_{ii}\hat{\mathbf{D}}_X^{-3}\frac{\partial \hat{\mathbf{A}}}{\partial s_{kl}} - \frac{1}{4}\delta_{kl}\delta_{cd}\delta_{kc}\mathbf{E}_{kk}\hat{\mathbf{D}}_X^{-3}\frac{\partial \hat{\mathbf{A}}}{\partial s_{ij}}\\
&\quad +\frac{\delta_{ij}}{2}\mathbf{E}_{ii}\hat{\mathbf{D}}_X^{-1}\frac{\partial^2 \hat{\mathbf{A}}}{\partial s_{kl}\,\partial s_{cd}} + \frac{\delta_{kl}}{2}\mathbf{E}_{kk}\hat{\mathbf{D}}_X^{-1}\frac{\partial^2 \hat{\mathbf{A}}}{\partial s_{ij}\,\partial s_{cd}}\\
&\quad +\frac{\delta_{cd}}{2}\mathbf{E}_{cc}\hat{\mathbf{D}}_X^{-1}\frac{\partial^2 \hat{\mathbf{A}}}{\partial s_{ij}\,\partial s_{kl}} + \hat{\mathbf{D}}_X\frac{\partial^3 \hat{\mathbf{A}}}{\partial s_{ij}\,\partial s_{kl}\,\partial s_{cd}}\\
&\quad (p+q \geqslant i \geqslant j \geqslant 1;\ p+q \geqslant k \geqslant l \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1).
\end{aligned} \tag{7.6}$$

The results for $\hat{\mathbf{B}}_{1\rho}$ are similarly obtained.

8. Issues on robustness and multiple roots

The asymptotic distributions of the parameter estimators in canonical correlation analysis under nonnormality are generally different from those under normality, which was illustrated in the numerical examples. In this section, we consider the robustness for the NT asymptotic distributions of sample canonical correlations using a model with latent variables. Let f_i and g_i be the ith elements of $\mathbf{f}$ and $\mathbf{g}$, respectively ($i = 1, \ldots, p$). Assume that

$$\begin{aligned}
&f_i = \lambda_i^{K_i}\tau_i + (1-\lambda_i^{2K_i})^{1/2}\varepsilon_{f_i}, \quad g_i = \lambda_i^{1-K_i}\tau_i + (1-\lambda_i^{2(1-K_i)})^{1/2}\varepsilon_{g_i},\\
&\mathrm{Var}(\tau_i) = \mathrm{Var}(\varepsilon_{f_i}) = \mathrm{Var}(\varepsilon_{g_i}) = 1, \quad 0 < K_i < 1 \ (i = 1, \ldots, p),
\end{aligned} \tag{8.1}$$

where τ_i, ε_{f_i} and ε_{g_i} are independently distributed. Then, it is easily seen that $\mathrm{Cor}(f_i, g_i) = \lambda_i$. A latent-variable model for $\mathbf{x}$ and $\mathbf{y}$ which gives (8.1) is shown in Section A.1 of the Appendix.

From now on, the subscript i is omitted for simplicity of notation. It is known that $\mathrm{avar}(\hat{\lambda})$ is the same as that of the usual sample correlation coefficient using the population coefficients of canonical variables i.e., the ith columns of $\mathbf{A}$ and $\mathbf{B}_1$ [30, Propositions 3 and 4]. Consequently, using the formula of the asymptotic variance for the sample correlation coefficient (e.g., [31, Eq. (3.4)]), we have

$$n\,\mathrm{avar}(\hat{\lambda}) = \rho_{ffgg} + \frac{\lambda^2}{4}(\rho_{ffff} + \rho_{gggg} + 2\rho_{ffgg}) - \lambda(\rho_{fffg} + \rho_{fggg}), \tag{8.2}$$

where $\rho_{abcd} = \sigma_{abcd}/(\sigma_{aa}\sigma_{bb}\sigma_{cc}\sigma_{dd})^{1/2}$.

When (8.1) holds with the independence assumption, we have

$$\rho_{ffff} = \lambda^{4K}\kappa_4(\tau) + (1-\lambda^{2K})^2\kappa_4(\varepsilon_f) + 3, \quad \rho_{fffg} = \lambda^{1+2K}\kappa_4(\tau) + 3\lambda,$$
$$\rho_{ffgg} = \lambda^2\kappa_4(\tau) + 2\lambda^2 + 1, \quad \rho_{fggg} = \lambda^{3-2K}\kappa_4(\tau) + 3\lambda,$$
$$\rho_{gggg} = \lambda^{4(1-K)}\kappa_4(\tau) + (1-\lambda^{2(1-K)})^2\kappa_4(\varepsilon_g) + 3. \tag{8.3}$$

From (8.2) and (8.3), it follows that

$$\begin{aligned} n\,\mathrm{avar}(\hat{\lambda}) &= \lambda^2\kappa_4(\tau) + 2\lambda^2 + 1 + \frac{\lambda^2}{4}\{(\lambda^{4K} + 2\lambda^2 + \lambda^{4(1-K)})\kappa_4(\tau) + (1-\lambda^{2K})^2\kappa_4(\varepsilon_f) \\ &\quad + (1-\lambda^{2(1-K)})^2\kappa_4(\varepsilon_g) + 4\lambda^2 + 8\} - \lambda\{(\lambda^{1+2K} + \lambda^{3-2K})\kappa_4(\tau) + 6\lambda\} \\ &= (1-\lambda^2)^2 + \left\{\lambda^2 + \frac{1}{4}(\lambda^{2+4K} + 2\lambda^4 + \lambda^{6-4K}) - (\lambda^{2+2K} + \lambda^{4-2K})\right\}\kappa_4(\tau) \\ &\quad + \frac{\lambda^2}{4}(1-\lambda^{2K})^2\kappa_4(\varepsilon_f) + \frac{\lambda^2}{4}(1-\lambda^{2(1-K)})^2\kappa_4(\varepsilon_g) \\ &\equiv (1-\lambda^2)^2 + \lambda_\tau\kappa_4(\tau) + \lambda_{\varepsilon_f}\kappa_4(\varepsilon_f) + \lambda_{\varepsilon_g}\kappa_4(\varepsilon_g), \end{aligned} \tag{8.4}$$

which gives

Theorem. *When the model of* (8.1) *with the mutual independence of* τ, ε_f *and* ε_g *holds, and when*

$$\lambda_\tau\kappa_4(\tau) + \lambda_{\varepsilon_f}\kappa_4(\varepsilon_f) + \lambda_{\varepsilon_g}\kappa_4(\varepsilon_g) = 0, \tag{8.5}$$

avar($\hat{\lambda}$) *under nonnormality is the same as that under normality.*

Note that the nonnormal cases with (8.5) exist, given arbitrary values of λ_τ, λ_{ε_f} and λ_{ε_g}, since the fourth cumulants take negative and positive values although there is no reason to expect that (8.5) will be satisfied in practice. Ogasawara [24, Eq. (5.4)]) gave the corresponding result in the case of the usual sample correlation coefficient with $K = \frac{1}{2}$. For this case, (8.5) becomes

$$\lambda^2(1-\lambda)^2\left\{\kappa_4(\tau) + \frac{\kappa_4(\varepsilon_f)}{4} + \frac{\kappa_4(\varepsilon_g)}{4}\right\} = 0. \tag{8.6}$$

That is, when $\kappa_4(\tau) + [\{\kappa_4(\varepsilon_f) + \kappa_4(\varepsilon_g)\}/4] = 0$, we have the robust NT avar($\hat{\lambda}$) irrespective of the values of λ. Eqs. (8.5) and (8.6) are seen as compensatory effects of the kurtosis of the latent variables.

Using the delta method with the results of Theorem gives

Corollary. *Under the same conditions as in Theorem for* $\hat{\lambda}_i$ $(i = 1, \ldots, p)$ *with the additional independence condition between different sets of latent variables* $[\tau_i, \varepsilon_{f_i}, \varepsilon_{g_i}]$ *and* $[\tau_j, \varepsilon_{f_j}, \varepsilon_{g_j}]$ $(i \neq j)$, *the NT asymptotic variances of order* $O(n^{-1})$ *for* $\hat{\lambda}_i^2$ $(i = 1, \ldots, p)$, $\hat{l}^*$, $\hat{\rho}_{XY}^2$ *and* $\hat{\rho}_{XY}$ *hold under nonnormality.*

For the NT asymptotic bias of the sample correlation coefficient, a similar result of the conditional robustness is available [24, Eq. (5.5)]. Unfortunately, the result does not convey to the cases of canonical correlations since the property corresponding to (8.2) required for the asymptotic bias does not hold.

Finally, we consider the problem of multiple roots or equal canonical correlations with some roots being 0. Let $\mathbf{\Lambda} = \mathrm{diag}[\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_{M^*}, \mathbf{0}']$, where $\mathbf{\Lambda}_i = \lambda_i \mathbf{I}_{m_i}$ $(i = 1, \ldots, M^*)$ with $1 > \lambda_1 > \cdots > \lambda_{M^*} > 0$, $p = \sum_{i=1}^{M^*} m_i + m_{M^*+1}$, and m_{M^*+1} is the number of 0 roots. Then, the covariance structure model of $\mathbf{\Sigma}$ in (2.6) is not identified when some of m_i is greater than 1. Assume that for some j $(j \leqslant M^*)$, $m_j \geqslant 2$. The corresponding two sets of m_j rows of $\mathbf{A}^{-1}$ and $[\mathbf{B}^{11}\ \mathbf{B}^{12}]$ are only identified up to the same orthogonal rotation. This indeterminacy can be removed by choosing some orthogonal rotation [6, p. 247] as was done in order to remove the rotational indeterminacy in $[\mathbf{B}^{21}\ \mathbf{B}^{22}]$ when $q \geqslant p + 2$. A simple example is to set appropriate $m_j(m_j - 1)/2$ elements in $\mathbf{A}^{-1}$ or $[\mathbf{B}^{11}\ \mathbf{B}^{12}]$ to 0. When $m_{M^*+1} \geqslant 1$, the four submatrices, $\mathbf{B}^{ij}$'s, should be reconstructed by the first $p - m_{M^*+1}$ rows/columns and the remaining $q - p + m_{M^*+1}$ ones in $\mathbf{B}^{-1}$, where the rotational indeterminacy, when $q - p + m_{M^*+1} \geqslant 2$, in the reconstructed $[\mathbf{B}^{21}\ \mathbf{B}^{22}]$ should be removed as before. When $m_{M^*+1} \geqslant 2$ the rotational indeterminacy occurs also for $\mathbf{A}^{-1}$, and should be removed as for $[\mathbf{B}^{21}\ \mathbf{B}^{22}]$. Redefining $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ after rotation, the asymptotic expansions in the previous sections can be similarly obtained, when m_j's are available.

Since the sample roots are distinct with probability 1, the covariance structure model with multiple roots is no longer a saturated one. Consequently, formulas (4.3)–(4.5) cannot be used. Instead, we require discrepancy functions for estimation of the parameters, which yields the estimators depending on the functions employed. The partial derivatives of $\hat{\boldsymbol{\theta}}$ with respect to $\mathbf{s}$ tend to become involved. The case of unweighted least squares estimation will be shown in Section A.2 of the appendix.

Appendix A.

A.1. An expression of canonical variables when the inter-batter factor analysis model holds

Let $\mathbf{x}$ and $\mathbf{y}$ in (2.1) be given by the latent variables in inter-battery factor analysis [33, 26, Section 8f; 7,34] as

$$\begin{aligned}
&\mathbf{x} = \boldsymbol{\mu}_X + \mathbf{C}_X \boldsymbol{\tau} + \boldsymbol{\Psi}_X^{1/2} \boldsymbol{\varepsilon}_X, \quad \boldsymbol{\Psi}_X^{1/2} \boldsymbol{\Psi}_X'^{1/2} = \boldsymbol{\Psi}_X, \quad \mathrm{Cov}(\boldsymbol{\tau}) = \mathrm{Cov}(\boldsymbol{\varepsilon}_X) = \mathbf{I}_p, \\
&\mathbf{y} = \boldsymbol{\mu}_Y + \mathbf{C}_Y \boldsymbol{\tau} + \boldsymbol{\Psi}_Y^{1/2} \boldsymbol{\varepsilon}_Y, \quad \boldsymbol{\Psi}_Y^{1/2} \boldsymbol{\Psi}_Y'^{1/2} = \boldsymbol{\Psi}_Y, \quad \mathrm{Cov}(\boldsymbol{\varepsilon}_Y) = \mathbf{I}_q, \\
&\mathrm{Cov}(\boldsymbol{\tau}, \boldsymbol{\varepsilon}_X') = \mathbf{O}, \quad \mathrm{Cov}(\boldsymbol{\tau}, \boldsymbol{\varepsilon}_Y') = \mathrm{Cov}(\boldsymbol{\varepsilon}_X, \boldsymbol{\varepsilon}_Y') = \mathbf{O},
\end{aligned} \tag{A.1}$$

where $\boldsymbol{\tau}$ is a $p \times 1$ vector of common or inter-battery factors; $\boldsymbol{\varepsilon}_X$ and $\boldsymbol{\varepsilon}_Y$ are $p \times 1$ and $q \times 1$ vectors of battery-specific factors, respectively; $\mathbf{C}_X$, $\mathbf{C}_Y$, $\boldsymbol{\Psi}_X^{1/2}$ and $\boldsymbol{\Psi}_Y^{1/2}$ are the loading matrices of $\mathbf{x}$ and $\mathbf{y}$ on the associated factors.

When canonical correlations are nonzero and distinct, from (2.6) and (A.1),

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{C}_X \mathbf{C}_X' + \boldsymbol{\Psi}_X & \mathbf{C}_X \mathbf{C}_Y' \\ \mathbf{C}_Y \mathbf{C}_X' & \mathbf{C}_Y \mathbf{C}_Y' + \boldsymbol{\Psi}_Y \end{bmatrix} = \begin{bmatrix} (\mathbf{A}\mathbf{A}')^{-1} & \mathbf{A}'^{-1} \mathbf{\Lambda} [\mathbf{B}^{11}\ \mathbf{B}^{12}] \\ \begin{bmatrix} \mathbf{B}'^{11} \\ \mathbf{B}'^{12} \end{bmatrix} \mathbf{\Lambda}\, \mathbf{A}^{-1} & (\mathbf{B}\mathbf{B}')^{-1} \end{bmatrix}. \tag{A.2}$$

From (A.2), $\mathbf{C}_X$, $\mathbf{C}_Y$, $\boldsymbol{\Psi}_X$ and $\boldsymbol{\Psi}_Y$ can be given by the parameters in canonical correlation analysis though there exists indeterminacy of factor transformation as in the usual factor analysis. It is known that the indeterminacy is more than the rotational one found in factor analysis. In this appendix, the following expressions of loadings are employed:

$$\mathbf{C}_X = \mathbf{A}'^{-1}\mathbf{\Lambda}^{(K)}, \quad \mathbf{C}_Y = \begin{bmatrix} \mathbf{B}'^{11} \\ \mathbf{B}'^{12} \end{bmatrix} \mathbf{\Lambda}^{(1-K)}, \tag{A.3}$$

where

$$\mathbf{\Lambda}^{(K)} = \mathrm{diag}(\lambda_1^{K_1}, \ldots, \lambda_p^{K_p}), \quad \mathbf{\Lambda}^{(1-K)} = \mathrm{diag}(\lambda_1^{1-K_1}, \ldots, \lambda_p^{1-K_p}),$$

$$1 > \lambda_1 > \cdots > \lambda_p > 0, \quad 0 < K_i < 1 \ \ (i = 1, \ldots, p).$$

The matrices $\mathbf{\Psi}_X$ and $\mathbf{\Psi}_Y$ are given by

$$\mathbf{\Psi}_X = (\mathbf{AA}')^{-1} - \mathbf{C}_X\mathbf{C}_X' = \mathbf{A}'^{-1}(\mathbf{I}_p - \mathbf{\Lambda}^{2(K)})\mathbf{A}^{-1} > \mathbf{O},$$

$$\begin{aligned}\mathbf{\Psi}_Y &= (\mathbf{BB}')^{-1} - \mathbf{C}_Y\mathbf{C}_Y' \\ &= \begin{bmatrix} \mathbf{B}'^{11} \\ \mathbf{B}'^{12} \end{bmatrix} (\mathbf{I}_p - \mathbf{\Lambda}^{2(1-K)})[\mathbf{B}^{11}\ \mathbf{B}^{12}] + \begin{bmatrix} \mathbf{B}'^{21} \\ \mathbf{B}'^{22} \end{bmatrix} [\mathbf{B}^{21}\ \mathbf{B}^{22}] > \mathbf{O},\end{aligned} \tag{A.4}$$

where $\mathbf{\Lambda}^{2(\cdot)} = (\mathbf{\Lambda}^{(\cdot)})^2$; and the inequalities are used in Löwner's sense. The positive definite property consistent with the latent variable model was first shown by Rao [26, Section 8f.3] when $K_1 = \cdots = K_p = \frac{1}{2}$.

Using (A.1) and (A.3) with $\mathbf{\Psi}_X^{1/2} = \mathbf{A}'^{-1}(\mathbf{I}_p - \mathbf{\Lambda}^{2(K)})^{1/2}$ and $\mathbf{\Psi}_Y^{1/2} = \mathbf{B}'^{-1}\begin{bmatrix} (\mathbf{I}_p - \mathbf{\Lambda}^{2(1-K)})^{1/2} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_q \end{bmatrix}$ from (A.4), it follows that

$$\begin{aligned}\mathbf{f} &= \mathbf{A}'(\mathbf{x} - \boldsymbol{\mu}_X) = \mathbf{A}'(\mathbf{A}'^{-1}\mathbf{\Lambda}^{(K)}\boldsymbol{\tau} + \mathbf{A}'^{-1}(\mathbf{I}_p - \mathbf{\Lambda}^{2(K)})^{1/2}\boldsymbol{\varepsilon}_X) \\ &= \mathbf{\Lambda}^{(K)}\boldsymbol{\tau} + (\mathbf{I}_p - \mathbf{\Lambda}^{2(K)})^{1/2}\boldsymbol{\varepsilon}_X,\end{aligned}$$

$$\begin{aligned}\mathbf{g} &= \mathbf{B}_1'(\mathbf{y} - \boldsymbol{\mu}_Y) = \mathbf{B}_1'\left(\begin{bmatrix} \mathbf{B}'^{11} \\ \mathbf{B}'^{12} \end{bmatrix} \mathbf{\Lambda}^{(1-K)}\boldsymbol{\tau} + \mathbf{B}'^{-1}\begin{bmatrix} (\mathbf{I}_p - \mathbf{\Lambda}^{2(1-K)})^{1/2} & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_q \end{bmatrix}\boldsymbol{\varepsilon}_Y\right) \\ &= \mathbf{\Lambda}^{(1-K)}\boldsymbol{\tau} + [(\mathbf{I}_p - \mathbf{\Lambda}^{2(1-K)})^{1/2}\ \mathbf{O}]\boldsymbol{\varepsilon}_Y = \mathbf{\Lambda}^{(1-K)}\boldsymbol{\tau} + (\mathbf{I}_p - \mathbf{\Lambda}^{2(1-K)})^{1/2}\boldsymbol{\varepsilon}_g,\end{aligned} \tag{A.5}$$

where $\boldsymbol{\varepsilon}_g$ is a vector whose elements are the first p elements of $\boldsymbol{\varepsilon}_Y$. It is seen that (A.5) with $\boldsymbol{\varepsilon}_X \equiv \boldsymbol{\varepsilon}_f = (\varepsilon_{f_1}, \ldots, \varepsilon_{f_p})'$ and $\boldsymbol{\varepsilon}_g = (\varepsilon_{g_1}, \ldots, \varepsilon_{g_p})'$ is equivalent to (8.1) when an additional assumption of the mutual independence among τ_i, ε_{f_i} and ε_{g_i} is employed.

A.2. The partial derivatives of the parameter estimators by unweighted least squares with respect to sample variances and covariances

Let $\hat{\boldsymbol{\theta}}$ be the $Q \times 1$ vector which minimizes

$$F_{\mathrm{LS}} = \tfrac{1}{2}\,\mathrm{tr}\{(\mathbf{\Sigma} - \mathbf{S})^2\}, \tag{A.6}$$

where $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$; $\boldsymbol{\theta}$ is a vector of mathematical variables with respect to which differentiation is performed while $\boldsymbol{\theta}$ is also used as a population vector for simplicity of notation; and Q is redefined as the number of nonduplicated free parameters in $\boldsymbol{\Sigma}$. Then, we have the first-order conditions for $\hat{\boldsymbol{\theta}}$

$$\frac{\partial \hat{F}_{\mathrm{LS}}}{\partial \hat{\theta}_i} \equiv \left.\frac{\partial F_{\mathrm{LS}}}{\partial \theta_i}\right|_{\theta_i=\hat{\theta}_i} = \mathrm{tr}\left\{(\hat{\boldsymbol{\Sigma}} - \mathbf{S})\frac{\partial \hat{\boldsymbol{\Sigma}}}{\partial \hat{\theta}_i}\right\} = 0 \quad (i = 1, \ldots, Q), \tag{A.7}$$

where $\partial\hat{\boldsymbol{\Sigma}}/\partial\hat{\theta}_i = \partial\boldsymbol{\Sigma}/\partial\theta_i|_{\theta_i=\hat{\theta}_i}$. The first partial derivatives are given by differentiating (A.7) with respect $\mathbf{s}$, which yields

$$\frac{\partial \hat{\boldsymbol{\theta}}}{\partial s_{ab}} = -\left(\frac{\partial^2 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial \hat{\boldsymbol{\theta}}'}\right)^{-1} \frac{\partial^2 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial s_{ab}} \quad (p+q \geqslant a \geqslant b \geqslant 1). \tag{A.8}$$

Differentiating (A.7) two and three times with respect to $\mathbf{s}$, we have

$$\begin{aligned}\frac{\partial^2 \hat{\boldsymbol{\theta}}}{\partial s_{ab}\,\partial s_{cd}} = &-\left(\frac{\partial^2 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial \hat{\boldsymbol{\theta}}'}\right)^{-1} \left(\sum_i \sum_j \frac{\partial^3 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial\hat{\theta}_i\partial\hat{\theta}_j}\frac{\partial\hat{\theta}_i}{\partial s_{ab}}\frac{\partial\hat{\theta}_j}{\partial s_{cd}} + \sum_i \frac{\partial^3 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial\hat{\theta}_i\,\partial s_{cd}}\frac{\partial\hat{\theta}_i}{\partial s_{ab}}\right.\\ &\left. + \sum_i \frac{\partial^3 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial\hat{\theta}_i\,\partial s_{ab}}\frac{\partial\hat{\theta}_i}{\partial s_{cd}} + \frac{\partial^3 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial s_{ab}\,\partial s_{cd}}\right)\end{aligned} \tag{A.9}$$

and

$$\begin{aligned}\frac{\partial^3 \hat{\boldsymbol{\theta}}}{\partial s_{ab}\,\partial s_{cd}\,\partial s_{ef}} = &-\left(\frac{\partial^2 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial \hat{\boldsymbol{\theta}}'}\right)^{-1} \left[\sum_i \sum_j \sum_k \frac{\partial^4 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial\hat{\theta}_i\,\partial\hat{\theta}_j\partial\hat{\theta}_k}\frac{\partial\hat{\theta}_i}{\partial s_{ab}}\frac{\partial\hat{\theta}_j}{\partial s_{cd}}\frac{\partial\hat{\theta}_k}{\partial s_{ef}}\right.\\ &+ \sum_{(U,V,W)}^{3} \sum_i \left\{\sum_j \left(\frac{\partial^3 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial\hat{\theta}_i\,\partial\hat{\theta}_j}\frac{\partial\hat{\theta}_i}{\partial s_U}\frac{\partial^2\hat{\theta}_j}{\partial s_V\partial s_W} + \frac{\partial^4 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial\hat{\theta}_i\,\partial\hat{\theta}_j\partial s_U}\frac{\partial\hat{\theta}_i}{\partial s_V}\frac{\partial\hat{\theta}_j}{\partial s_W}\right)\right.\\ &\left.\left. + \frac{\partial^3 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial\hat{\theta}_i\,\partial s_U}\frac{\partial^2\hat{\theta}_i}{\partial s_V\partial s_W} + \frac{\partial^4 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial\hat{\theta}_i\,\partial s_U\partial s_V}\frac{\partial\hat{\theta}_i}{\partial s_W}\right\} + \frac{\partial^4 \hat{F}_{\mathrm{LS}}}{\partial \hat{\boldsymbol{\theta}}\,\partial s_{ab}\,\partial s_{cd}\,\partial s_{ef}}\right]\\ &(p+q \geqslant a \geqslant b \geqslant 1;\ p+q \geqslant c \geqslant d \geqslant 1;\ p+q \geqslant e \geqslant f \geqslant 1),\end{aligned} \tag{A.10}$$

where $\sum_i$ is $\sum_{i=1}^{Q}$ and $\sum_{(U,V,W)}^{3}$ denotes a summation over the range $(U, V, W) \in \{(ab, cd, ef), (cd, ef, ab), (ef, ab, cd)\}$.

The second partial derivatives of $\hat{\boldsymbol{\Sigma}}$ with respect to $\hat{\boldsymbol{\theta}}$ and $\mathbf{s}$ evaluated at the population values are

$$\begin{aligned}&\frac{\partial^2 F_{\mathrm{LS}}}{\partial\theta_i\,\partial\theta_j} = \mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right), \quad \left.\frac{\partial^2 \hat{F}_{\mathrm{LS}}}{\partial\hat{\theta}_i\partial s_{cd}}\right|_{\mathbf{s}=\boldsymbol{\sigma}} = -(2-\delta_{cd})\frac{\partial\sigma_{cd}}{\partial\theta_i}\\ &(i, j = 1, \ldots, Q;\ p+q \geqslant c \geqslant d \geqslant 1).\end{aligned} \tag{A.11}$$

Similarly, the third partial derivatives are

$$\begin{aligned}&\frac{\partial^3 F_{\mathrm{LS}}}{\partial\theta_i\,\partial\theta_j\,\partial\theta_k} = \mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_j\,\partial\theta_k} + \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\,\partial\theta_k} + \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\,\partial\theta_j}\right),\\ &\left.\frac{\partial^3 \hat{F}_{\mathrm{LS}}}{\partial\hat{\theta}_i\,\partial\hat{\theta}_j\,\partial s_{cd}}\right|_{\mathbf{s}=\boldsymbol{\sigma}} = -(2-\delta_{cd})\frac{\partial^2\sigma_{cd}}{\partial\theta_i\,\partial\theta_j} \quad (i, j, k = 1, \ldots, Q;\ p+q \geqslant c \geqslant d \geqslant 1).\end{aligned} \tag{A.12}$$

The fourth partial derivatives are given by

$$\frac{\partial^4 F_{\mathrm{LS}}}{\partial\theta_i\,\partial\theta_j\,\partial\theta_k\partial\theta_l} = \mathrm{tr}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\frac{\partial^3\boldsymbol{\Sigma}}{\partial\theta_j\partial\theta_k\partial\theta_l} + \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\frac{\partial^3\boldsymbol{\Sigma}}{\partial\theta_i\,\partial\theta_k\partial\theta_l} + \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\frac{\partial^3\boldsymbol{\Sigma}}{\partial\theta_i\,\partial\theta_j\partial\theta_l}\right.$$
$$\left. + \frac{\partial\boldsymbol{\Sigma}}{\partial\theta_l}\frac{\partial^3\boldsymbol{\Sigma}}{\partial\theta_i\,\partial\theta_j\partial\theta_k} + \frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\,\partial\theta_j}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_k\partial\theta_l} + \frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\partial\theta_k}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_j\,\partial\theta_l} + \frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_i\,\partial\theta_l}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\theta_j\partial\theta_k}\right),$$

$$\left.\frac{\partial^4 \hat{F}_{\mathrm{LS}}}{\partial\hat{\theta}_i\,\partial\hat{\theta}_j\,\partial\hat{\theta}_k\partial s_{cd}}\right|_{\mathbf{s}=\boldsymbol{\sigma}} = -(2-\delta_{cd})\frac{\partial^3\sigma_{cd}}{\partial\theta_i\,\partial\theta_j\partial\theta_k}$$
$$(i,j,k,l = 1,\ldots,Q;\ \ p+q \geqslant c \geqslant d \geqslant 1). \tag{A.13}$$

References

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, New York, 1958.

[2] T.W. Anderson, Asymptotic theory for canonical correlation analysis, J. Multivariate Anal. 70 (1999) 1–29.

[3] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, third ed., Wiley, New York, 2003.

[4] Z.D. Bai, X. He, A chi-square test for dimensionality with non-Gaussian data, J. Multivariate Anal. 88 (2004) 109–117.

[5] M.S. Bartlett, Further aspects of the theory of multiple regression, Proc. Cambridge Philos. Soc. 34 (1938) 33–40.

[6] R.J. Boik, A local parameterization of orthogonal and semi-orthogonal matrices with applications, J. Multivariate Anal. 67 (1998) 244–276.

[7] M.W. Browne, The maximum likelihood solution in inter-battery factor analysis, British J. Math. Statist. Psychology 32 (1979) 75–86.

[8] M.W. Browne, Asymptotically distribution-free methods for the analysis of covariance structures, British J. Math. Statist. Psychology 37 (1984) 62–83.

[9] A.G. Constantine, Some non-central distribution problems in multivariate analysis, Ann. Math. Statist. 34 (1963) 1270–1285.

[10] G.S. Eaton, E. Ronchetti, General saddlepoint approximations with applications to L statistics, J. Amer. Statist. Assoc. 81 (1986) 420–430.

[11] M.L. Eaton, D. Tyler, The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis, J. Multivariate Anal. 50 (1994) 238–264.

[12] C. Fang, P.R. Krishnaiah, Asymptotic distributions of functions of the eigenvalues of some random matrices for nonnormal populations, J. Multivariate Anal. 12 (1982) 39–63.

[13] Y. Fujikoshi, Asymptotic expansions for the distributions of some multivariate tests, in: P.R. Krishnaiah (Ed.), Multivariate Analysis-IV, North-Holland, New York, 1977, pp. 55–71.

[14] Y. Fujikoshi, Asymptotic expansions for the distributions of the sample roots under nonnormality, Biometrika 67 (1980) 45–51.

[15] P. Hall, On the removal of skewness by transformation, J. Roy. Statist. Soc. B 54 (1992) 221–228.

[16] P. Hall, The Bootstrap and Edgeworth Expansion, Springer, New York, 1992, corrected printing, 1997.

[17] H. Hotelling, The most predictable criterion, J. Educational Psychology 26 (1935) 139–142.

[18] H. Hotelling, Relations between two sets of variates, Biometrika 28 (1936) 321–377.

[19] P.L. Hsu, On the limiting distribution of the canonical correlations, Biometrika 32 (1941) 38–45.

[20] R. H Koning, H. Neudecker, T. Wansbeek, Unbiased estimation of fourth-order matrix moments, Linear Algebra Appl. 160 (1992) 163–174.

[21] D.N. Lawley, Tests of significance in canonical analysis, Biometrika 46 (1959) 59–66.

[22] D.N. Lawley, A.E. Maxwell, Factor Analysis as a Statistical Method, second ed., Butterworths, London, 1971.

[23] R.J. Muirhead, C.M. Waternaux, Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, Biometrika 67 (1980) 31–43.

[24] H. Ogasawara, Asymptotic expansion of the sample correlation coefficient under nonnormality, Comput. Statist. Data Anal. 50 (2006) 891–910.

[25] H. Ogasawara, Asymptotic expansion of the distributions of the estimators in factor analysis under nonnormality, British J. Math. Statist. Psychology 2007, to appear.
[26] C.R. Rao, Linear Statistical Inference and its Applications, second ed., Wiley, New York, 1973.
[27] A.C. Rencher, Methods of Multivariate Analysis, second ed., Wiley, New York, 2002.
[28] W.W. Rozeboom, Linear correlations between sets of variables, Psychometrika 30 (1965) 57–71.
[29] M. Siotani, T. Hayakawa, Y. Fujikoshi, Modern Multivariate Statistical Analysis: A Graduate Course and Handbook, American Sciences Press, Columbus, OH, 1985.
[30] J.H. Steiger, M.W. Browne, The comparison of interdependent correlations between optimal linear composite, Psychometrika 49 (1984) 11–24.
[31] J.H. Steiger, A.R. Hakstian, The asymptotic distribution of elements of a correlation matrix: theory and application, British J. Math. Statist. Psychology 35 (1982) 208–215.
[32] T. Tsuchiya, S. Konishi, General saddlepoint approximations and normalizing transformations for multivariate statistics, Commun. Statist. Theory Methods 26 (1997) 2541–2563.
[33] L.R. Tucker, An inter-battery method of factor analysis, Psychometrika 23 (1958) 111–136.
[34] J.A. Wegelin, A. Packer, T.S. Richardson, Latent models for cross-covariance, J. Multivariate Anal. 97 (2006) 79–102.